Numerically stable online estimation of variance in particle filters

JIMMY OLSSON[1] and RANDAL DOUC[2]

[1]*Department of Mathematics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden.*
E-mail: jimmyol@kth.se
[2]*Département CITI, TELECOM SudParis, 9 rue Charles Fourier, 91000 EVRY, France.*
E-mail: randal.douc@it-sudparis.eu

This paper discusses variance estimation in sequential Monte Carlo methods, alternatively termed particle filters. The variance estimator that we propose is a natural modification of that suggested by H.P. Chan and T.L. Lai [*Ann. Statist.* **41** (2013) 2877–2904], which allows the variance to be estimated in a single run of the particle filter by tracing the genealogical history of the particles. However, due particle lineage degeneracy, the estimator of the mentioned work becomes numerically unstable as the number of sequential particle updates increases. Thus, by tracing only a part of the particles' genealogy rather than the full one, our estimator gains long-term numerical stability at the cost of a bias. The scope of the genealogical tracing is regulated by a lag, and under mild, easily checked model assumptions, we prove that the bias tends to zero geometrically fast as the lag increases. As confirmed by our numerical results, this allows the bias to be tightly controlled also for moderate particle sample sizes.

Keywords: asymptotic variance; Feynman–Kac models; hidden Markov models; particle filters; sequential Monte Carlo methods; state-space models; variance estimation

1. Introduction

Since the *bootstrap particle filter* was introduced in [14], *sequential Monte Carlo* (SMC) *methods*, alternatively termed *particle filters*, have been successfully applied within a wide range of applications, including computer vision, automatic control, signal processing, optimisation, robotics, econometrics, and finance; see, for example, [13,23] for introductions to the topic. SMC methods approximate a given sequence of distributions by a sequence of possibly weighted empirical measures associated with a sample of *particles* evolving recursively and randomly in time. Each iteration of the SMC algorithm comprises two operations: a *mutation step*, which moves the particles randomly in the state space, and a *selection step*, which duplicates/eliminates, through resampling, particles with high/low importance weights, respectively.

In parallel with algorithmic developments, the theoretical properties of SMC have been studied extensively during the last twenty years, and there is currently a number of available results describing the convergence, as the number of particles tends to infinity, of Monte Carlo estimates produced by the algorithm; see, for example, the monographs [5,6] and [1], Chapter 9. The first *central limit theorem* (CLT) for SMC methods was established in [7], and this result was later refined in the series of papers [3,10,18]. In the mentioned CLT, the asymptotic variance of the weak Gaussian limit is expressed through a recursive formula involving high-dimensional integrals over generally complicated integrands and is hence intractable in general.

Due to its complexity, only a very few recent works have treated the important—although challenging—topic of variance estimation in SMC algorithms. A breakthrough was made by H.P. Chan and T.L. Lai, who proposed, in [2] and within the framework of general *state-space models* (or, *hidden Markov models*), an estimator, from now on referred to as the *Chan & Lai estimator* (CLE), that allows the sequence of asymptotic variances to be estimated on the basis of a *single* realisation of the algorithm, without the need of additional simulations. Remarkably, the CLE can be shown to be consistent (see [2], Theorem 2), that is, to converge in probability to the true asymptotic variance as the number of particles tends to infinity. At a given time step, the CLE estimates the asymptotic variance by tracing genealogically the time-zero ancestors of the particles at the time step in question. The variance estimators proposed recently in [19] are based on the same principle, and may be viewed as refinements of the CLE within a more general framework of *Feynman–Kac models* and particle algorithms with time varying particle population sizes. Moreover, [19] provides an elegant, deepened (asymptotic as well as non-asymptotic) theoretical analysis of the technique, and these results are essential for the development of the present paper.

Appealingly, the set of time-zero ancestors may be updated recursively in the SMC algorithm by adding just a single line to the code. This allows variance estimates to be computed online with essentially the same computational complexity and memory demands as the original algorithm. Nevertheless, since the SMC algorithm performs repeatedly selection, it is well established that all the particles in the sample will, eventually, share the *same* time-zero ancestor (see, e.g., [16] for a theoretical analysis of this particle path degeneracy phenomenon). Unfortunately, this implies that the CLE collapses eventually to zero as time increases. Increasing the particle sample size or resampling less frequently the particles will postpone somewhat, but not avoid, this collapse. This makes the CLE impractical in the long run. Thus, although the SMC methodology has become a standard tool in statistics and engineering, a numerically stable estimator of the SMC variance has, surprisingly, hitherto been lacking, and the aim of the present paper is to—at least partially—fill this gap.

The natural solution that we propose in the present paper is to estimate the SMC asymptotic variance by tracing only a *part* of the particles' genealogy rather than the full one (i.e., back to the time-zero ancestors). By tracing genealogically only the last generations, depleted ancestor sets are avoided as long as the particle sample size is at least moderately large. Still, this measure leads to a bias, whose size determines completely the success of the approach. Nevertheless, in [12], which studies the stochastic stability of the sequence of asymptotic variances within the framework of general hidden Markov models, or, viewed differently, randomly perturbed Feynman–Kac models (see [12], Remark 1), it is established that the variance, which at time n may be expressed in the form of a sum of $n + 1$ terms, is, in the case where the perturbations form a stationary sequence, uniformly stochastically bounded in time, or, *tight*. Moreover, the analysis provided in [12], which is driven by mixing assumptions, indicates that the size of the mth term in the variance at time n decreases geometrically fast with the difference $n - m$. Consequently, we may expect the last terms of the sum to represent the major part of the variance, and as long as the number λ of traced generations, the *lag*, and the particle sample size are not too small, the bias should be negligible. This argument is confirmed by our main result, Theorem 6, which at any time point n provides an order ρ^λ bound on the asymptotic (as the number of particles tends to infinity) bias, where $\rho \in (0, 1)$ is a mixing rate. Consequently, as long as the number of

particles is large enough, the bias stays numerically stable in the long run and may, as it decreases geometrically fast with the lag λ, be controlled efficiently. Methodologically, the estimator that we propose has similarities with the *fixed-lag smoothing* approach studied in [17,21,22], and we here face the same bias-variance tradeoff, in the sense that a too greedy/generous lag design leads to high bias/variance, respectively.

The developments of the present paper are cast into the framework of randomly perturbed Feynman–Kac models, and the theoretical analysis is driven by the assumptions of [11,12] (going back to [9]), which are easily checked for many models used in practice. In particular, by replacing the now classical strong mixing assumption on the transition kernel of the underlying Markov chain—a standard assumption in the literature that typically requires the state space of the Markov chain to be a compact set—by a *local Doeblin condition*, we are able to verify the assumptions for a wide class of models with possibly non-compact state space.

As a numerical illustration, we apply our estimator in the context of SMC-based predictor flow approximation in general state-space models, including the widely used *stochastic volatility* model proposed in [15]. We are able to report an efficient, numerically stable performance of our variance estimator, with tight control of the bias at low variance.

The paper is structured as follows. Section 2 introduces some notation, defines the framework of perturbed Feynman–Kac models, and provides some background to SMC. In Section 3, focus is set on asymptotic variance estimation and after a prefatory discussion on the CLE we introduce the proposed fixed-lag variance estimator. We also describe how our estimator can be straightforwardly extended to so-called *updated* Feynman–Kac distribution flows. Theoretical and numerical results are found in Section 4 and Section 5, respectively, and the Appendix contains all proofs. Our theoretical analysis is divided into two parts: first, the identification of the limiting bias and, second, the construction of a tight upper bound on the same. The first part relies on the theoretical machinery developed in [19], whose key elements are recalled briefly in the beginning of Section A.1. Finally, Section 6 concludes the paper.

2. Preliminaries

2.1. Some notation and conventions

We assume that all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The set of natural numbers is denoted by $\mathbb{N} = \{0, 1, 2, \ldots\}$, and we let $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ be the positive ones. For all $(m, n) \in \mathbb{N}^2$, we set $[\![m, n]\!] := \{m, m + 1, \ldots, n\}$. The set of nonnegative real numbers is denoted by $\mathbb{R}_+$. For any quantities $\{a_\ell\}_{\ell=1}^m$, vectors are denoted by $\bar{a}_\ell^m := (a_\ell, \ldots, a_m)$.

We introduce some measure and kernel notation. Given some state space $(\mathsf{E}, \mathcal{E})$, we denote by $\mathbb{F}(\mathcal{E})$ and $\mathbb{M}(\mathcal{E})$ the spaces of bounded measurable functions and probability measures on $(\mathsf{E}, \mathcal{E})$, respectively. For any functions $(h, h') \in \mathbb{F}(\mathcal{E})^2$ we define the product function $h \otimes h' :$ $\mathsf{E}^2 \ni (x, x') \mapsto h(x)h'(x')$. The identity function $x \mapsto x$ is denoted by id. Let μ be a measure on $(\mathsf{E}, \mathcal{E})$; then for any μ-integrable function h, we denote by

$$\mu h := \int h(x)\mu(\mathrm{d}x)$$

the Lebesgue integral of h w.r.t. μ. In addition, let $(\mathsf{E}', \mathcal{E}')$ be some other measurable space and $\mathbf{K}$ some possibly unnormalised transition kernel $\mathbf{K} : \mathsf{E} \times \mathcal{E}' \to \mathbb{R}_+$. The kernel $\mathbf{K}$ induces two integral operators, one acting on functions and the other on measures. More specifically, given a measure ν on $(\mathsf{E}, \mathcal{E})$ and a measurable function h on $(\mathsf{E}', \mathcal{E}')$, we define the measure

$$\nu \mathbf{K} : \mathcal{E}' \ni A \mapsto \int \mathbf{K}(x, A)\nu(\mathrm{d}x)$$

and the function

$$\mathbf{K}h : \mathsf{E} \ni x \mapsto \int h(y)\mathbf{K}(x, \mathrm{d}y),$$

whenever these quantities are well defined.

2.2. Randomly perturbed Feynman–Kac models

Let $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Z}, \mathcal{Z})$ be a pair of general measurable spaces. Moreover, let $\mathbf{K}$ and χ be a Markov transition kernel and a probability measure on $(\mathsf{X}, \mathcal{X})$, respectively, and $\{g\langle z\rangle : z \in \mathsf{Z}\}$ a family of real-valued, positive, and measurable *potential functions* on $(\mathsf{X}, \mathcal{X})$. For all vectors $\bar{z}_k^m \in \mathsf{Z}^{m-k+1}$, we define unnormalised transition kernels

$$\mathbf{Q}\langle \bar{z}_k^m \rangle : \mathsf{X} \times \mathcal{X} \ni (x_k, A) \mapsto \int \cdots \int \mathbb{1}_A(x_{m+1}) \prod_{\ell=k}^{m} g\langle z_\ell\rangle(x_\ell)\mathbf{M}(x_\ell, \mathrm{d}x_{\ell+1}),$$

with the convention $\mathbf{Q}\langle \bar{z}_k^m \rangle(x, A) = \delta_x(A)$ if $m < k$ (where δ_x denotes the Dirac mass located at x), and probability measures

$$\eta\langle \bar{z}_k^m \rangle : \mathcal{X} \ni A \mapsto \frac{\chi \mathbf{Q}\langle \bar{z}_k^m \rangle \mathbb{1}_A}{\chi \mathbf{Q}\langle \bar{z}_k^m \rangle \mathbb{1}_\mathsf{X}}. \tag{1}$$

Using these definitions we may, given a sequence $\{z_n\}_{n \in \mathbb{N}}$ of *perturbations* in Z, express the *Feynman–Kac distribution flow* $\{\eta\langle \bar{z}_0^n \rangle\}_{n \in \mathbb{N}}$ recursively as

$$\eta\langle \bar{z}_0^n \rangle = \frac{\eta\langle \bar{z}_0^{n-1} \rangle \mathbf{Q}\langle z_n \rangle}{\eta\langle \bar{z}_0^{n-1} \rangle \mathbf{Q}\langle z_n \rangle \mathbb{1}_\mathsf{X}}, \qquad n \in \mathbb{N} \tag{2}$$

(where, by the previous convention, $\eta\langle \bar{z}_0^{-1} \rangle = \chi$). Even though the previous model may be applied in a non-temporal context, we will often refer to the index n as "time".

Example 1 (Partially dominated state-space models). Let $(\mathsf{X}, \mathcal{X})$ be a measurable space, $\mathbf{M} : \mathsf{X} \times \mathcal{X} \to [0, 1]$ a Markov transition kernel, and χ a probability measure on $(\mathsf{X}, \mathcal{X})$ (the latter being referred to as the *initial distribution*). In addition, let $(\mathsf{Y}, \mathcal{Y})$ be another measurable space and $g : \mathsf{X} \times \mathsf{Y} \to \mathbb{R}_+$ a Markov transition density with respect to some reference measure ν on

$(\mathsf{Y}, \mathcal{Y})$. By a general state-space model, we mean the canonical version of the bivariate Markov chain $\{(X_n, Y_n)\}_{n \in \mathbb{N}}$ having transition kernel

$$\mathsf{X} \times \mathsf{Y} \times \mathcal{X} \otimes \mathcal{Y} \ni ((x, y), A) \mapsto \iint \mathbb{1}_A(x', y') g(x', y') \nu(\mathrm{d}y') \mathbf{M}(x, \mathrm{d}x')$$

and initial distribution

$$\mathcal{X} \otimes \mathcal{Y} \ni A \mapsto \iint \mathbb{1}_A(x, y) g(x, y) \nu(\mathrm{d}y) \chi(\mathrm{d}x).$$

Here the marginal process $\{X_n\}_{n \in \mathbb{N}}$, referred to as the *state process*, is only partially observed through the *observation process* $\{Y_n\}_{n \in \mathbb{N}}$. For the model $\{(X_n, Y_n)\}_{n \in \mathbb{N}}$ defined in this way,

   (i)   the state process is a Markov chain with transition kernel $\mathbf{M}$ and initial distribution χ,
   (ii) the observations are, given the states, conditionally independent and such that the marginal conditional distribution of each Y_n depends on X_n only and has density $g(X_n, \cdot)$

(we refer to [1], Section 2.2 for details). When operating on a well-specified state-space model, a key ingredient is typically the computation of the flow of *predictor distributions*, where the predictor $\eta\langle \bar{y}_0^{n-1} \rangle$ at time $n \in \mathbb{N}$ is defined as the conditional distribution of the state X_n given the record $\bar{y}_0^{n-1} \in \mathsf{Y}^n$ of realised historical observations up to time $n - 1$. Using Bayes' formula (see, e.g., [1], Section 3.2.2 for details), it is straightforwardly shown that the predictor flow satisfies a perturbed Feynman–Kac recursion (2) with $(\mathsf{X}, \mathcal{X})$, $\mathbf{M}$, and χ given above, the observations $\{Y_n\}_{n \in \mathbb{N}}$ playing the role of perturbations (i.e., $\mathsf{Z} \leftarrow \mathsf{Y}$ and $\mathcal{Z} \leftarrow \mathcal{Y}$), and the local likelihood functions $\{g(\cdot, y) : y \in \mathsf{Y}\}$ playing the role of potential functions $\{g\langle y \rangle : y \in \mathsf{Y}\}$. We will return to this framework in Section 5.

2.3. Sequential Monte Carlo methods

SMC methods approximate online the Feynman–Kac flow generated by (2) and a given sequence $\{z_n\}_{n \in \mathbb{N}}$ of perturbations by propagating recursively a random sample $\{\xi_n^i\}_{i=1}^N$ of X-valued *particles*. More specifically, given a particle sample $\{\xi_n^i\}_{i=1}^N$ *targeting* $\eta\langle \bar{z}_0^{n-1} \rangle$ in the sense that for all $h \in \mathbb{F}(\mathcal{X})$, $\eta_N\langle \bar{z}_0^{n-1} \rangle h \simeq \eta\langle \bar{z}_0^{n-1} \rangle h$ as N tends to infinity, where

$$\eta_N\langle \bar{z}_0^{n-1} \rangle : \mathcal{X} \ni A \mapsto \frac{1}{N} \sum_{i=1}^N \mathbb{1}_A(\xi_n^i)$$

denotes the empirical measure associated with the particles, an updated particle sample $\{\xi_{n+1}^i\}_{i=1}^N$ approximating $\eta\langle \bar{z}_0^n \rangle$ is, as the perturbation z_n becomes accessible, formed by Algorithm 1.

   (In the algorithm above, $\mathsf{Cat}(\{\omega_n^\ell / \Omega_n\}_{\ell=1}^N)$ denotes the categorical distribution induced by the normalised particle weights $\{\omega_n^\ell / \Omega_n\}_{\ell=1}^N$.) Algorithm 1 is initialised at time $n = 0$ by drawing $\{\xi_0^i\}_{i=1}^N \sim \chi^{\otimes N}$. For all $n \in \mathbb{N}$ and all $h \in \mathbb{F}(\mathcal{X})$, the convergence, as N tends to infinity, of $\eta_N\langle \bar{z}_0^{n-1} \rangle h$ to $\eta\langle \bar{z}_0^{n-1} \rangle h$ can be established in several probabilistic senses. In particular, the first

Data: $\{\xi_n^i\}_{i=1}^N$, z_n
Result: $\{\xi_{n+1}^i\}_{i=1}^N$
1 set $\Omega_n \leftarrow 0$;
2 **for** $i = 1 \rightarrow N$ **do**
3   set $\omega_n^i \leftarrow g\langle z_n\rangle(\xi_n^i)$;
4   set $\Omega_n \leftarrow \Omega_n + \omega_n^i$;
5 **for** $i = 1 \rightarrow N$ **do**
6   draw $I_{n+1}^i \sim \mathsf{Cat}(\{\omega_n^\ell / \Omega_n\}_{\ell=1}^N)$;
7   draw $\xi_{n+1}^i \sim \mathbf{M}(\xi_n^{I_{n+1}^i}, \cdot)$;

Algorithm 1: SMC particle update

CLT for SMC methods was provided by [7], establishing that

$$\sqrt{N}\big(\eta_N\langle\bar{z}_0^{n-1}\rangle h - \eta\langle\bar{z}_0^{n-1}\rangle h\big) \xrightarrow{\mathcal{D}} \sigma\langle\bar{z}_0^{n-1}\rangle(h)Z, \tag{3}$$

where Z is standard normally distributed and the asymptotic variance is given by $\sigma\langle\bar{z}_0^{n-1}\rangle :=$ $\sigma_0\langle\bar{z}_0^{n-1}\rangle$ with

$$\sigma_\ell^2\langle\bar{z}_0^{n-1}\rangle : \mathbb{F}(\mathcal{X}) \ni h \mapsto \sum_{m=\ell}^n \frac{\eta\langle\bar{z}_0^{m-1}\rangle\{\mathbf{Q}\langle\bar{z}_m^{n-1}\rangle(h - \eta\langle\bar{z}_0^{n-1}\rangle h)\}^2}{(\eta\langle\bar{z}_0^{m-1}\rangle\mathbf{Q}\langle\bar{z}_m^{n-1}\rangle\mathbb{1}_{\mathsf{x}})^2} \tag{4}$$

(see also [3,10,18] for similar results). The fact that we in (4) define a truncated version of the variance with only $n - \ell + 1$ terms will be clear later on. In the coming section, we propose a lag-based, numerically stable estimator of the sequence $\{\sigma^2\langle\bar{z}_0^{n-1}\rangle\}_{n\in\mathbb{N}}$ of asymptotic variances. The estimator approximates $\{\sigma^2\langle\bar{z}_0^{n-1}\rangle\}_{n\in\mathbb{N}}$ online, as n increases, under constant computational complexity and memory requirements. Importantly, the estimator is obtained as a by-product of the particle filter output and does not require additional simulations. The numerical stability is obtained at the price of a small bias, which may be controlled under weak assumptions on the mixing properties of the model.

3. A lag-based variance estimator

3.1. The variance estimator proposed in [2]

Since Algorithm 1 resamples the particles at each time step, the particle cloud may be associated with a tree describing the genealogical lineages of the particles. The estimators proposed in [2] and [19] are based on the particles' *Eve indices* $\{E_n^i\}_{i=1}^N$ (the terminology is adopted from [19]), which are, for all $n \in \mathbb{N}$, defined as the indices of the time-zero ancestors of the particles $\{\xi_n^i\}_{i=1}^N$. More specifically, the Eve indices may, for all $i \in [\![1, N]\!]$, be computed recursively in Algorithm 1

(just after Line 6) by letting

$$E_n^i := \begin{cases} i & \text{for } n = 0, \\ E_{n-1}^{I_n^i} & \text{for } n \in \mathbb{N}^*. \end{cases}$$

Using the Eve indices, H.P. Chan and T.L. Lai proposed, in [2], for all $n \in \mathbb{N}$, $\sigma_N^2 \langle \bar{z}_0^{n-1} \rangle (h)$, with

$$\sigma_N^2 \langle \bar{z}_0^{n-1} \rangle : \mathbb{F}(\mathcal{X}) \ni h \mapsto \frac{1}{N} \sum_{i=1}^N \left(\sum_{j : E_n^j = i} \left\{ h(\xi_n^j) - \eta_N \langle \bar{z}_0^{n-1} \rangle h \right\} \right)^2, \tag{5}$$

as an estimator of $\sigma^2 \langle \bar{z}_0^{n-1} \rangle (h)$ for all $h \in \mathbb{F}(\mathcal{X})$. As mentioned in the introduction, we will refer to this estimator as the CLE. (More precisely, in [2], focus was set on the *updated* distribution flows discussed in Section 3.3 below; the adaptation is however straightforward.) In [19], a generalisation of the CLE, allowing the particle population size N to vary between SMC iterations, is presented. As the main result of [2], the consistency, as N tends to infinity, of the CLE is established; see also [19], Theorem 1 and Corollary 1 for a generalisation.

The CLE is indeed remarkable, as it allows the variance to be estimated online in a single run of the particle filter with no further simulation. Nevertheless, as explained in the introduction, the previous estimator has a serious flaw which is related to the well-known *particle path depletion phenomenon* of SMC algorithms. More specifically, resampling the particles systematically at each time leads without exception to a random time point before which all the genealogical traces coincide; we refer again to [16], which provides a time uniform $\mathcal{O}(N \log N)$ bound on the expected number of generations back in time to this most recent common ancestor. Thus, as n increases, the sets $\{j \in [\![1, N]\!] : E_n^j = i\}$ will eventually be empty for all indices $i \in [\![1, N]\!]$ except one, say, i_0, for which $\{j \in [\![1, N]\!] : E_n^j = i_0\} = [\![1, N]\!]$. As a consequence, eventually, $\sigma_N^2 \langle \bar{z}_0^{n-1} \rangle (h) = 0$ for all $h \in \mathbb{F}(\mathcal{X})$, which makes the estimator impractical. In the next section, we propose a simple modification of the CLE that stabilises numerically the same at the cost of a negligible, controllable bias.

3.2. Our estimator

The estimator that we propose is based on the simple idea of stabilising numerically the CLE by tracing, backwards in time, only a few generations of the particle genealogy, rather than tracing the history all the way back to the time-zero ancestors. In our approach, the Eve indices will be replaced by *Enoch indices*[1] defined, for all $i \in [\![1, N]\!]$ and $m \in \mathbb{N}$, recursively as

$$E_{m,n}^i := \begin{cases} i & \text{for } n = m, \\ E_{m,n-1}^{I_n^i} & \text{for } n > m. \end{cases} \tag{6}$$

In other words, for all $n \in \mathbb{N}$, $m \in [\![1, n]\!]$, and $i \in [\![1, N]\!]$, $\xi_m^{E_{m,n}^i}$ is the ancestor of ξ_n^i at time m. Now, let $\lambda \in \mathbb{N}$ be some fixed number, referred to as the lag, and define $n(\lambda) := (n - \lambda) \vee 0$; then,

---

[1] Two figures named Enoch appear in the 2nd as well as the 6th generations of the Genealogies of Genesis, as the son of Cain and the son-son-son-son-son of Seth, respectively.

Data: $\{\xi_n^i\}_{i=1}^N$, $\{E_{n(\lambda),n}^i, \ldots, E_{n,n}^i\}_{i=1}^N$, z_n
Result: $\{\xi_{n+1}^i\}_{i=1}^N$, $\{E_{(n+1)(\lambda),n+1}^i, \ldots, E_{n+1,n+1}^i\}_{i=1}^N$

1 set $\Omega_n \leftarrow 0$;
2 **for** $i = 1 \to N$ **do**
3   set $\omega_n^i \leftarrow g\langle z_n\rangle(\xi_n^i)$;
4   set $\Omega_n \leftarrow \Omega_n + \omega_n^i$;
5 **for** $i = 1 \to N$ **do**
6   draw $I_{n+1}^i \sim \mathsf{Cat}(\{\omega_n^\ell / \Omega_n\}_{\ell=1}^N)$;
7   draw $\xi_{n+1}^i \sim \mathbf{M}(\xi_n^{I_{n+1}^i}, \cdot)$;
8   **for** $m = (n+1)(\lambda) \to n$ **do**
9   set $E_{m,n+1}^i \leftarrow E_{m,n}^{I_{n+1}^i}$;
10   set $E_{n+1,n+1}^i \leftarrow i$;

Algorithm 2: SMC particle and Enoch-index update

we propose $\sigma_{N,\lambda}^2\langle \bar{z}_0^{n-1}\rangle(h)$, with

$$\sigma_{N,\lambda}^2\langle \bar{z}_0^{n-1}\rangle : \mathbb{F}(\mathcal{X}) \ni h \mapsto \frac{1}{N} \sum_{i=1}^N \left(\sum_{j:E_{n(\lambda),n}^j=i} \{h(\xi_n^j) - \eta_N\langle \bar{z}_0^{n-1}\rangle h\} \right)^2, \quad (7)$$

as an estimator of the variance $\sigma^2\langle \bar{z}_0^{n-1}\rangle(h)$ for all $n \in \mathbb{N}$, $\bar{z}_0^{n-1} \in \mathsf{Z}^n$, and $h \in \mathbb{F}(\mathcal{X})$. On-line computation of the Enoch indices $\{E_{n(\lambda),n}^i\}_{i=1}^N$ requires the propagation of a window $\{E_{n(\lambda),n}^i, \ldots, E_{n,n}^i\}_{i=1}^N$ of indices; see Algorithm 2 for a pseudo-code. As the length of the window is bounded by $\lambda + 1$, the memory demand of the estimator is $\mathcal{O}(\lambda N)$ independently of n. Moreover, since genealogical tracing has a linear complexity in N, the total complexity of the estimator is $\mathcal{O}(\lambda N)$, again independently of n.

For $n = 0$, Algorithm 2 is initialised by drawing $\{\xi_0^i\}_{i=1}^N \sim \chi^{\otimes N}$ and setting $E_{0,0}^i \leftarrow i$ for all $i \in [\![1, N]\!]$. At the end of the algorithm, after the second **for**-loop, an estimate

$$\sigma_{N,\lambda}^2\langle \bar{z}_0^n\rangle(h) = \frac{1}{N} \sum_{i=1}^N \left(\sum_{j:E_{(n+1)(\lambda),n+1}^j=i} \{h(\xi_{n+1}^j) - \eta_N\langle \bar{z}_0^n\rangle h\} \right)^2$$

of $\sigma_\lambda^2\langle \bar{z}_0^n\rangle(h)$ may be formed for all $h \in \mathbb{F}(\mathcal{X})$.

3.3. Variance estimators for flows of updated distributions

Some applications involve approximation of the *updated* measures

$$\bar{\eta}\langle \bar{z}_k^m\rangle : \mathcal{X} \ni A \mapsto \frac{\chi\mathbf{Q}\langle \bar{z}_k^{m-1}\rangle(g\langle z_m\rangle \mathbb{1}_A)}{\chi\mathbf{Q}\langle \bar{z}_k^{m-1}\rangle(g\langle z_m\rangle \mathbb{1}_{\mathsf{X}})}, \quad (8)$$

for $\bar{z}_k^m \in \mathsf{Z}^{m-k+1}$, rather than the measures defined by (1).

Example 2 (Partially dominated state-space models, revisited). In the case of the partially dominated state-space models discussed in Example 1, the updated measures $\{\bar{\eta}\langle\bar{y}_0^n\rangle\}_{n\in\mathbb{N}}$ defined through (8) are the *filter distributions*; more precisely, in this context, for all $n \in \mathbb{N}$, $\bar{\eta}\langle\bar{y}_0^n\rangle$ is the conditional distribution of the state X_n given the realised observations $\bar{y}_0^n \in \mathsf{Y}^{n+1}$ up to time n (i.e., *including* the last observation y_n).

Since for all $h \in \mathbb{F}(\mathcal{X})$, by normalisation,

$$\bar{\eta}\langle\bar{z}_k^m\rangle h = \frac{\eta\langle\bar{z}_k^{m-1}\rangle(g\langle z_m\rangle h)}{\eta\langle\bar{z}_k^{m-1}\rangle g\langle z_m\rangle},$$

the flow $\{\bar{\eta}\langle\bar{z}_0^n\rangle\}_{n\in\mathbb{N}}$ of updated distributions is naturally approximated by the flow of weighted empirical measures

$$\bar{\eta}_N\langle\bar{z}_0^n\rangle : A \ni \mathcal{X} \mapsto \frac{\eta_N\langle\bar{z}_k^{m-1}\rangle(g\langle z_m\rangle\mathbb{1}_A)}{\eta_N\langle\bar{z}_k^{m-1}\rangle g\langle z_m\rangle} = \sum_{i=1}^N \frac{\omega_n^i}{\Omega_n}\mathbb{1}_A(\xi_n^i), \tag{9}$$

for some given sequence $\{z_n\}_{n\in\mathbb{N}}$ of perturbations, where the weights $\{\omega_n^i\}_{i=1}^N$ and the weight sum Ω_n are computed in Algorithm 1. By the normality (3) and the consistency (43) one obtains, using Slutsky's theorem, for all $\bar{z}_0^n \in \mathsf{Z}^{n+1}$, the central limit theorem

$$\sqrt{N}\big(\bar{\eta}\langle\bar{z}_0^n\rangle h - \bar{\eta}\langle\bar{z}_0^n\rangle h\big) \xrightarrow{\mathcal{D}} \bar{\sigma}\langle\bar{z}_0^n\rangle(h)Z, \tag{10}$$

as N tends to infinity, where Z is standard normally distributed and the asymptotic variance is given by $\bar{\sigma}^2\langle\bar{z}_0^n\rangle(h) = \bar{\sigma}_0^2\langle\bar{z}_0^n\rangle(h)$ with

$$\bar{\sigma}_\ell^2\langle\bar{z}_0^n\rangle : \mathbb{F}(\mathcal{X}) \ni h \mapsto \frac{\sigma_\ell^2\langle\bar{z}_0^{n-1}\rangle(g\langle z_n\rangle\{h - \bar{\eta}\langle\bar{z}_0^n\rangle h\})}{(\eta\langle\bar{z}_0^{n-1}\rangle g\langle z_n\rangle)^2} \tag{11}$$

(where $\sigma_\ell\langle\bar{z}_0^{n-1}\rangle$ is defined in (4) for the original Feynman–Kac particle model). In the case $\ell = 0$, the expression (11) is found also in [12], Eqn. (17). In the light of (11), casting our fixed-lag approach into the framework of updated Feynman–Kac models yields the estimator

$$\bar{\sigma}_{N,\lambda}^2\langle\bar{z}_0^n\rangle : \mathbb{F}(\mathcal{X}) \ni h \mapsto \frac{\sigma_{N,\lambda}^2\langle\bar{z}_0^{n-1}\rangle(g\langle z_n\rangle\{h - \bar{\eta}_N\langle\bar{z}_0^n\rangle h\})}{(\eta_N\langle\bar{z}_0^{n-1}\rangle g\langle z_n\rangle)^2}$$

$$= N \sum_{i=1}^N \left(\sum_{j:E_{n(\lambda),n}^j=i} \frac{\omega_n^i}{\Omega_n}\{h(\xi_n^j) - \bar{\eta}_N\langle\bar{z}_0^n\rangle h\}\right)^2 \tag{12}$$

for some suitable lag $\lambda \in \mathbb{N}$ (where the equality stems from the fact that $\eta_N\langle\bar{z}_0^{n-1}\rangle\{g\langle z_n\rangle(h - \bar{\eta}_N\langle\bar{z}_0^n\rangle h)\} = 0$).

4. Theoretical results

4.1. Main assumptions

In the following, we assess the theoretical properties of our estimator. All proofs are presented in the Appendix. We preface our main assumptions by the definition of an *r-local Doeblin set*.

Definition 3. Let $r \in \mathbb{N}^*$. A set $C \in \mathcal{X}$ is *r-local Doeblin* with respect to $\mathbf{M}$ and g if there exist positive functions $\varepsilon_C^- : \mathsf{Z}^r \to \mathbb{R}^+$ and $\varepsilon_C^+ : \mathsf{Z}^r \to \mathbb{R}^+$, a family $\{\mu_C \langle \bar{z} \rangle; \bar{z} \in \mathsf{Z}^r\}$ of probability measures, and a family $\{\varphi_C \langle \bar{z} \rangle; \bar{z} \in \mathsf{Z}^r\}$ of positive functions such that for all $\bar{z} \in \mathsf{Z}^r$, $\mu_C \langle \bar{z} \rangle(C) = 1$ and for all $A \in \mathcal{X}$ and $x \in C$,

$$\varepsilon_C^- \langle \bar{z} \rangle \varphi_C \langle \bar{z} \rangle(x) \mu_C \langle \bar{z} \rangle(A) \leq \mathbf{Q} \langle \bar{z} \rangle(x, A \cap C) \leq \varepsilon_C^+ \langle \bar{z} \rangle \varphi_C \langle \bar{z} \rangle(x) \mu_C \langle \bar{z} \rangle(A).$$

Our theoretical analysis is driven by the following assumptions, under which [12] established that the asymptotic SMC variance is stochastically bounded (tight) in time.

(A1). The perturbation process $\{\zeta_n\}_{n \in \mathbb{Z}}$, taking on values in $(\mathsf{Z}, \mathcal{Z})$, is strictly stationary and ergodic.[2] Moreover, there exist an integer $r \in \mathbb{N}^*$ and a set $K \in \mathcal{Z}^{\otimes r}$ such that the following holds.

 (i) The process $\{\bar{\zeta}_n\}_{n \in \mathbb{Z}}$, where $\bar{\zeta}_n := \bar{\zeta}_{nr}^{(n+1)r-1}$, is ergodic and such that $\mathbb{P}(\bar{\zeta}_0 \in K) > 2/3$.
 (ii) For all $\eta > 0$ there exists an r-local Doeblin set $C \in \mathcal{X}$ such that for all $\bar{z}_0^{r-1} \in K$,

$$\sup_{x \in C^{\complement}} \mathbf{Q} \langle \bar{z}_0^{r-1} \rangle \mathbb{1}_{\mathsf{X}}(x) \leq \eta \sup_{x \in \mathsf{X}} \mathbf{Q} \langle \bar{z}_0^{r-1} \rangle \mathbb{1}_{\mathsf{X}}(x) < \infty$$

 and

$$\inf_{\bar{z}_0^{r-1} \in K} \frac{\varepsilon_C^- \langle \bar{z}_0^{r-1} \rangle}{\varepsilon_C^+ \langle \bar{z}_0^{r-1} \rangle} > 0,$$

 where the functions ε_C^+ and ε_C^- are given in Definition 3.
 (iii) There exists a set $D \in \mathcal{X}$ such that

$$\mathbb{E}\left[\ln^- \inf_{x \in D} \mathbf{Q} \langle \bar{\zeta}_0^{r-1} \rangle \mathbb{1}_D(x)\right] < \infty.$$

(A2).

 (i) For all $(x, z) \in \mathsf{X} \times \mathsf{Z}$, $g \langle z \rangle(x) > 0$.
 (ii) For all $z \in \mathsf{Z}$, $\|g \langle z \rangle\|_\infty < \infty$.
 (iii) $\mathbb{E}[\ln^+ \|g \langle \zeta_0 \rangle\|_\infty] < \infty$.

---

[2]Here ergodic means, more specifically, that the dynamical system induced on $(\mathsf{Z}^{\mathbb{N}}, \mathcal{Z}^{\otimes \mathbb{N}})$ by the distribution of $\{\zeta_n\}_{n \in \mathbb{Z}}$ and the shift operator is ergodic.

Remark 4. In the case $r = 1$, **(A1)** may be replaced by the simpler assumption that there exists a set $K \in \mathcal{Z}$ such that the following holds.

   (i) $\mathbb{P}(\zeta_0 \in K) > 2/3$.
   (ii) For all $\eta > 0$ there exists a 1-local Doeblin set $C \in \mathcal{X}$ such that for all $z \in K$,

$$\sup_{x \in C^{\complement}} g\langle z\rangle(x) \leq \eta \|g\langle z\rangle\|_\infty < \infty. \tag{13}$$

  (iii) There exists a set $D \in \mathcal{X}$ satisfying

$$\inf_{x \in D} \mathbf{M}(x, D) > 0 \quad \text{and} \quad \mathbb{E}\left[\ln^- \inf_{x \in D} g\langle \zeta_0\rangle(x)\right] < \infty.$$

This simplified version will be used in Section 5.

4.2. Theoretical properties of the fixed-lag estimator

As established by the following proposition, for all $n \in \mathbb{N}$, $\bar{z}_0^{n-1} \in \mathsf{Z}^n$, and $\lambda \in \mathbb{N}$, the estimator $\sigma_{N,\lambda}^2\langle \bar{z}_0^{n-1}\rangle$ is a consistent for $\sigma_{n(\lambda)}^2\langle \bar{z}_0^{n-1}\rangle$, where the latter is given by (4) with $\ell = n(\lambda)$.

Proposition 5. *Assume* **(A2)**(i–ii). *Then for all* $n \in \mathbb{N}$, $\bar{z}_0^{n-1} \in \mathsf{Z}^n$, $\lambda \in \mathbb{N}$, *and* $h \in \mathbb{F}(\mathcal{X})$, *as* $N \to \infty$,

$$\sigma_{N,\lambda}^2\langle \bar{z}_0^{n-1}\rangle(h) \xrightarrow{\mathbb{P}} \sigma_{n(\lambda)}^2\langle \bar{z}_0^{n-1}\rangle(h).$$

Now, define, for all $n \in \mathbb{N}$ and $\bar{z}_0^{n-1} \in \mathsf{Z}^n$, the *asymptotic bias*

$$\beta_\lambda\langle \bar{z}_0^{n-1}\rangle : \mathbb{F}(\mathcal{X}) \ni h \mapsto \sigma^2\langle \bar{z}_0^{n-1}\rangle(h) - \sigma_{n(\lambda)}^2\langle \bar{z}_0^{n-1}\rangle(h), \tag{14}$$

which is always nonnegative. For the integer $r \in \mathbb{N}^*$ and the set $D \in \mathcal{X}$ given in **(A1)**, introduce the class

$$\mathbb{M}(D, r) := \left\{ \chi \in \mathbb{M}(\mathcal{X}) : \mathbb{E}\left[\ln^- \chi \mathbf{Q}\langle \bar{\zeta}_0^{\ell-1}\rangle \mathbb{1}_D\right] < \infty \text{ for all } \ell \in [\![0, r]\!] \right\} \tag{15}$$

of initial distributions. Then the following theorem, which is the main result of this paper, states that the asymptotic bias can be controlled uniformly in time by the lag λ. In particular, the bias decreases with λ at a geometric rate.

Theorem 6. *Assume* **(A1)**–**(A2)**. *Then there exist a constant* $\rho \in (0, 1)$ *and a sequence* $(c_k)_{k \in \mathbb{N}^*}$ *in* $\mathbb{R}_+$ *such that for all* $\chi \in \mathbb{M}(D, r)$ *(with* D *and* r *given by* **(A1)**), $\lambda \in \mathbb{N}$, $h \in \mathbb{F}(\mathcal{X})$, *and* $k \in \mathbb{N}$,

$$\sup_{n \in \mathbb{N}} \mathbb{P}\left(\frac{\beta_\lambda\langle \bar{\zeta}_0^{n-1}\rangle(h)}{\rho^{\lambda+1}\|h\|_\infty^2} > c_k \right) \leq \frac{1}{k}. \tag{16}$$

Remark 7. We stress that the sequence $\{c_k\}_{k\in\mathbb{N}^*}$ as well as the mixing rate ρ in the previous theorem depend exclusively on the properties of the model and not on the lag size λ, the objective function h, or the initial distribution χ (as long as this belongs to $\mathbb{M}(D, r)$). This means that at any level specified by k, the asymptotic bias can be controlled uniformly in time at a geometric rate in λ.

The strength of these bounds will be illustrated numerically in Section 5.

4.3. Extensions to variance estimators for flows of updated distributions

In the following, we confirm that the previous results can be extended to flows of updated Feynman–Kac distributions (recall the definitions in Section 3.3).

Proposition 8. *Assume* (A2)(i–ii). *Then for all* $n \in \mathbb{N}$, $\bar{z}_0^n \in \mathsf{Z}^{n+1}$, $\lambda \in \mathbb{N}$, *and* $h \in \mathbb{F}(\mathcal{X})$, *as* $N \to \infty$,

$$\bar{\sigma}_{N,\lambda}^2 \langle \bar{z}_0^n \rangle (h) \xrightarrow{\mathbb{P}} \bar{\sigma}_{n(\lambda)}^2 \langle \bar{z}_0^n \rangle (h).$$

Now, as in the previous section, define, for all $n \in \mathbb{N}$ and $\bar{z}_0^n \in \mathsf{Z}^{n+1}$, the asymptotic bias

$$\bar{\beta}_\lambda \langle \bar{z}_0^n \rangle : \mathbb{F}(\mathcal{X}) \ni h \mapsto \bar{\sigma}^2 \langle \bar{z}_0^n \rangle (h) - \bar{\sigma}_{n(\lambda)}^2 \langle \bar{z}_0^n \rangle (h). \tag{17}$$

Theorem 9. *Assume* (A1)–(A2). *Then the statement of Theorem 6 holds true when* $\beta_\lambda \langle \bar{\zeta}_0^{n-1} \rangle$ *is replaced by* $\bar{\beta}_\lambda \langle \bar{\zeta}_0^n \rangle$.

4.4. Bounds on the asymptotic bias under strong mixing assumptions

Before continuing to the numerical part of the paper, we provide, for the sake of completeness, a stronger bound on the asymptotic bias under the following *strong mixing* assumption, which is standard in the literature of SMC analysis (see [8] and, e.g., [1,4,5] for refinements) and points to applications where the state space X is a compact set.

(S).

(i) There exist constants $0 < \underline{\varepsilon} < \bar{\varepsilon} < \infty$ and a probability measure $\mu \in \mathbb{M}(\mathcal{X})$ such that for all $x \in \mathsf{X}$ and $A \in \mathcal{X}$,

$$\underline{\varepsilon}\mu(A) \le \mathbf{M}(x, A) \le \bar{\varepsilon}\mu(A).$$

(ii) There exist constants $0 < \underline{\delta} < \bar{\delta} < \infty$ such that for all $z \in \mathsf{Z}$, $\|g\langle z\rangle\|_\infty \le \bar{\delta}$ and for all $(x, z) \in \mathsf{X} \times \mathsf{Z}$, $\mathbf{M}g\langle z\rangle(x) \ge \underline{\delta}$.

Under (S)(i), define $\varrho := 1 - \underline{\varepsilon}/\bar{\varepsilon}$; then, instead of bounding the asymptotic bias stochastically (as in Theorem 6), the following theorem provides a *deterministic* bound on the same. The proof is found in Section A.5.

Theorem 10. *Assume* (**S**)*. Then for all $n \in \mathbb{N}$, $\bar{z}_0^{n-1} \in \mathbb{Z}^n$, $\lambda \in \mathbb{N}$, and $h \in \mathbb{F}(\mathcal{X})$,*

$$\beta_\lambda \langle \bar{z}_0^{n-1} \rangle (h) \leq \begin{cases} 0 & \text{for } n \leq \lambda, \\ c\varrho^{2(\lambda+1)} \|h\|_\infty^2 & \text{for } n > \lambda, \end{cases}$$

where

$$c := \frac{(\bar{\delta}/\underline{\delta})^2}{(1+\varrho)(1-\varrho)^3}.$$

Remark 11. Note that in the previous theorem, the strong mixing assumption (**S**) renders a bound on the asymptotic bias that is completely explicit in ϱ and λ. This allows us to derive a criterion for how the lag λ should be designed in order to guarantee a given bound $\varepsilon > 0$ on the asymptotic bias. More specifically, note that

$$c\varrho^{2(\lambda+1)} \|h\|_\infty^2 \leq \varepsilon \Rightarrow \lambda \geq \frac{\log(2\varepsilon) - 2\log(\|h\|_\infty \bar{\delta}/\underline{\delta}) + 3\log(1-\varrho)}{2\log \varrho} - 1. \tag{18}$$

The function $\varrho \mapsto \log(1-\varrho)/\log \varrho$, which plays the crucial role in the lower bound (18) on λ, is displayed in Figure 1. After a very slow increase over, say, $(0, 0.95)$, the function increases very
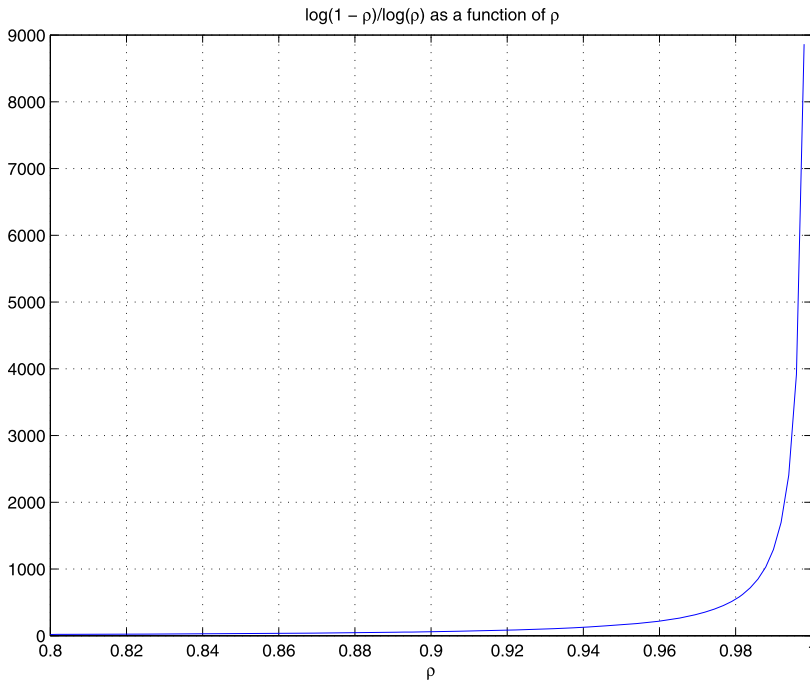

Figure 1. The function $\varrho \mapsto \log(1-\varrho)/\log \varrho$ over the interval $(0.8, 1)$.

rapidly as ϱ is approaching 1. We may hence expect the performance of the lag-based variance estimator to be relatively robust vis-á-vis the forgetting, in the sense that the model has to exhibit very poor mixing properties in order for the lag approximation to deteriorate. However, in the latter case the lag-based approximation introduces a significant bias, which is perfectly in line with our expectations.

5. Application to state-space models

5.1. Predicting log-volatility

As a first numerical illustration, we consider the problem of computing the predictor flow in the—now classical—stochastic volatility model

$$
\begin{aligned}
X_{n+1} &= \varphi X_n + \sigma U_{n+1}, \\
Y_n &= \beta \exp(X_n/2) V_n,
\end{aligned}
\qquad n \in \mathbb{N},
\tag{19}
$$

proposed in [15], where $\{U_n\}_{n \in \mathbb{N}^*}$ and $\{V_n\}_{n \in \mathbb{N}}$ are sequences of uncorrelated standard Gaussian noise. In this model, where only the process $\{Y_n\}_{n \in \mathbb{N}}$ is observable, β is a constant scaling factor, φ is the *persistence*, and σ is the *volatility of the log-volatility*. In the case where $|\varphi| < 1$, the state process $\{X_n\}_{n \in \mathbb{N}}$ is stationary with a Gaussian invariant distribution having zero mean and variance $\tilde{\sigma}^2 := \sigma^2/(1 - \varphi^2)$.

It is easily checked that the model $\{(X_n, Y_n)\}_{n \in \mathbb{N}}$ defined by (19) is indeed a state-space model in the notion of Example 1, and in this case $\mathsf{X} = \mathsf{Y} = \mathbb{R}$, $\mathcal{X} = \mathcal{Y} = \mathcal{B}(\mathbb{R})$, and

$$
\mathbf{M}(x, A) = \int_A \phi(x'; \varphi x, \sigma^2) \, dx',
$$

$$
g(x, y) = \phi(y; 0, \beta^2 e^x),
$$

where $(x, y) \in \mathbb{R}^2$, $A \in \mathcal{B}(\mathbb{R})$, and $\phi(\cdot; \mu, \varsigma^2)$ denotes the density of the Gaussian distribution with expectation $\mu \in \mathbb{R}$ and standard deviation $\varsigma > 0$. Consequently, the flow of one-step log-volatility predictors satisfies the perturbed Feynman–Kac recursion (2) with $g\langle y \rangle = g(\cdot, y)$, $y \in \mathsf{Y}$, and may thus be approximated using SMC methods.

We check that the model satisfies the simplified assumptions in Remark 4 for the scenario where $|\varphi| < 1$, implying stationary state and observation processes, the latter with marginal stationary distribution

$$
\mathcal{B}(\mathbb{R}) \ni A \mapsto \iint \mathbb{1}_A(y) \phi(y; 0, \beta^2 e^x) \phi(x; 0, \tilde{\sigma}^2) \, dx \, dy.
\tag{20}
$$

For this purpose, we first note that for all $y \in \mathbb{R}$,

$$
\| g\langle y \rangle \|_\infty = \frac{1}{|y|\sqrt{2\pi}} e^{-1/2},
$$

that is, for all $(x, y) \in \mathbb{R}^2$,

$$\frac{g\langle y\rangle(x)}{\|g\langle y\rangle\|_\infty} \propto |y| \exp\left(-\frac{x}{2} - \frac{y^2 e^{-x}}{2\beta^2}\right),$$

where the right hand side tends to zero as $|x| \to \infty$ for all $y \in \mathbb{R}$. We thus conclude that Conditions (i) and (ii) in Remark 4 hold true for any compact set $K \subset \mathbb{R}$ with probability exceeding $2/3$ under (20). In this case, every compact set $C \subset \mathbb{R}$ is 1-local Doeblin with respect to Lebesgue measure. In addition, since

$$\mathbb{E}[Y_0^2] = \beta^2 \int e^x \phi(x; 0, \tilde{\sigma}^2)\, dx < \infty, \tag{21}$$

Condition (iii) is satisfied for all compact sets $D \subset \mathbb{R}$. Moreover, for such compact sets D, (21) implies that $\mathbb{M}(D, 1)$ contains all initial distributions χ with $\chi(D) > 0$; in particular, $\mathbb{M}(D, 1)$ contains the invariant Gaussian distribution of the state process, which we use as initial distribution in the predictor Feynman–Kac recursion.

In this context, we aim at computing the sequence $\{\eta\langle \bar{y}_0^{n-1}\rangle \,\mathrm{id}\}_{n\in\mathbb{N}}$ of predictor means; however, due to the nonlinearity of the emission density g, closed-form expressions are beyond reach, and we thus apply SMC for this purpose. In the scenario considered by us, $(\beta, \varphi, \sigma) = (0.641, 0.975, 0.165)$, which are the parameter estimates obtained in [1], Example 11.1.1.2 on the basis of British pound/US dollar daily log-returns from 1 October 1981 to 28 June 1985. In this setting, the following two numerical experiments were conducted.

5.1.1. *Lag size influence*

First, in order to assess the dependence of the bias (14) on the lag λ, a record comprising 600 observations were generated by simulation of the model under the parameters above. We re-emphasise that even though we here, for simplicity, consider the scenario of a perfectly specified model, this is not required in our assumptions in Section 4 (as long as the observation sequence is stationary). By inputting, 100 times, these observations into Algorithm 2 and running the same with $N = 4000$ particles, 100 replicates of the particle predictor mean at time 600 were produced and furnished with estimates of the asymptotic variance at the same time point. For each replicate, the asymptotic variance was estimated using all the lags $\lambda \in \{2, 10, 12, 14, 16, 18, 20, 22, 50, 100, 200, 600\}$, where the last one, $\lambda = 600$, corresponds to the CLE. Moreover, the reference value 1.63 of the asymptotic variance, again at the time point 600, was obtained by the brute-force approach consisting in generating as many as 1000 replicates of the particle predictor mean, again with $N = 4000$ particles, and simply multiplying the sample variance of these replicates by N. The outcome is reported numerically and visually in Table 1 and Figure 2, respectively, where the different boxes in Figure 2 correspond to different lags. For each box, the reference value is indicated by a blue-colored asterisk ($*$) and the average estimate of each box is indicated by a red ditto. From Figure 2, it is evident that the clear bias introduced by the smallest lags ($\lambda \in \{2, 10, 12\}$) decreases rapidly as λ increases, and for $\lambda = 20$ the bias is more or less eliminated. After this, increasing further λ leads, as we may expect from the particle path degeneracy, to significant increase of variance and no further improvement of

Table 1. Means and standard deviations of the variance estimates reported in Figure 2

λ	Mean	St. dev.
2	0.47	0.02
10	0.96	0.09
12	0.99	0.11
14	1.47	0.40
16	1.54	0.49
18	1.60	0.57
20	1.63	0.62
22	1.62	0.61
50	1.58	0.61
100	1.53	0.71
200	1.46	0.71
600	1.30	0.96

the bias. On the contrary, the fact that the cardinality of the set $\{E^i_{n(\lambda),n} : i \in [\![1, N]\!]\}$ decreases monotonously as λ increases, leading to constantly zero variance estimates for n and λ large enough, re-introduces bias also for large λ. For the last replicate, the cardinalities of the sets $\{E^i_n : i \in [\![1, N]\!]\}$ and $\{E^i_{n(20),n} : i \in [\![1, N]\!]\}$ were 8 and 140, respectively.

5.1.2. *Long-term stability*

In order to investigate numerically the long-term stability of our fixed-lag estimator, we executed Algorithm 2 on a considerably longer observation record comprising 3500 values generated by simulation. The number of particles was set to $N = 5000$. Guided by the outcome of the previous experiment, we furnished the estimated predictor means with variance estimates obtained in the same sweep of the data using the fixed-lag estimator with $\lambda = 20$. In parallel, the CLE was computed on the same realisation of the particle cloud. Finally, the brute-force approach estimating the asymptotic variances on the basis of 1200 replicates of the predictor mean sequence was re-applied as a reference.

Figure 3 displays the time evolution of the variance estimates over the initial 200 time steps, where only estimates corresponding to every second time step have been plotted for readability. As evident from the top panel, the CLE targets well the reference values at most time points for this relatively limited time horizon, even though slight numerical instability may be discerned towards the very end of the plot. In addition, the fixed-lag estimator follows nicely the reference values for most time points with a variance that is somewhat smaller than that of the CLE. In order to display the estimators' long-term stability properties, Figure 4 provides the analogous plot for the full observation record comprising 3500 values. Again, for readability, only estimates corresponding to every 35th time step have been plotted. Now, as clear from the top panel, the estimates produced by the CLE degenerate rapidly and after, say, 1500 time steps the CLE loses track completely of the reference values. From time step 2871, all particles in the cloud share the same Eve index, and the CLE collapses to zero. On the the contrary, from the bottom panel
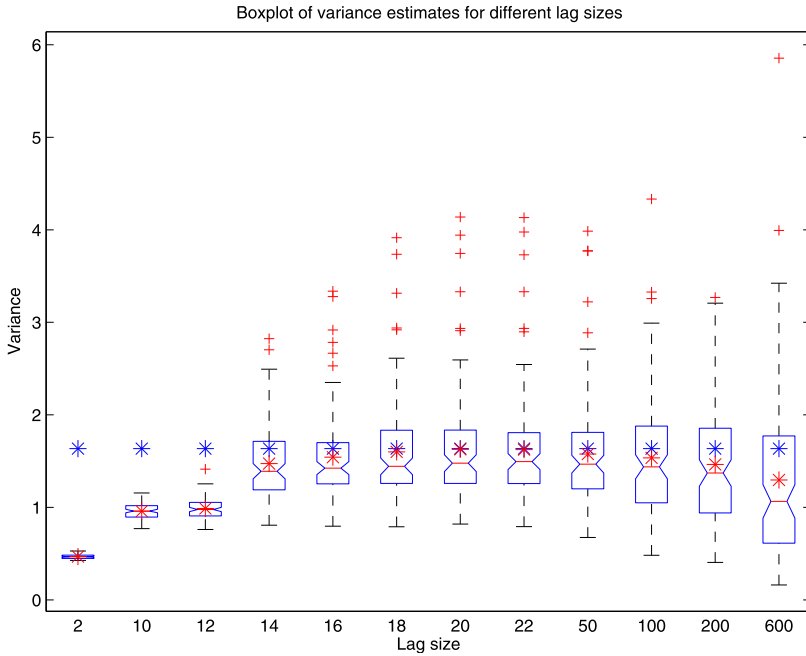
Figure 2. Estimated asymptotic variances of the particle predictor mean at time 600 in the stochastic volatility model (19). The particle population size is $N = 4000$. The boxes are based on 100 replicates of Algorithm 2 and correspond to the different lags $\lambda \in \{2, 10, 12, 14, 16, 18, 20, 22, 50, 100, 200, 600\}$. The reference value 1.63, which is indicated by blue-colored asterisks ($*$), was obtained as the sample variance (scaled by the number of particles) of 1000 independent replicates of the particle predictor mean at time 600 (again, Algorithm 2 used $N = 4000$ particles). Red-colored asterisks indicate average estimates.

it is evident that the estimates delivered by the fixed-lag estimator stay numerically stable and follows well the reference values at most time points. In particular, the variance peak arising as a result of extreme state process behavior at time 3395 is captured strikingly well by the fixed-lag estimator.

5.2. SMC confidence bounds

The *linear Gaussian state-space model*

$$
\begin{aligned}
X_{n+1} &= \varphi X_n + \sigma_u U_{n+1}, \\
Y_n &= X_n + \sigma_v V_n,
\end{aligned}
\qquad n \in \mathbb{N}, \tag{22}
$$

where $\{U_n\}_{n\in\mathbb{N}^*}$ and $\{V_n\}_{n\in\mathbb{N}}$ are again sequences of uncorrelated standard Gaussian noise, allows predictor means to be computed in a closed form using *Kalman prediction* (see, e.g., [1], Algorithm 5.2.9). Thus, allowing comparisons with true quantities to be made, the class of linear
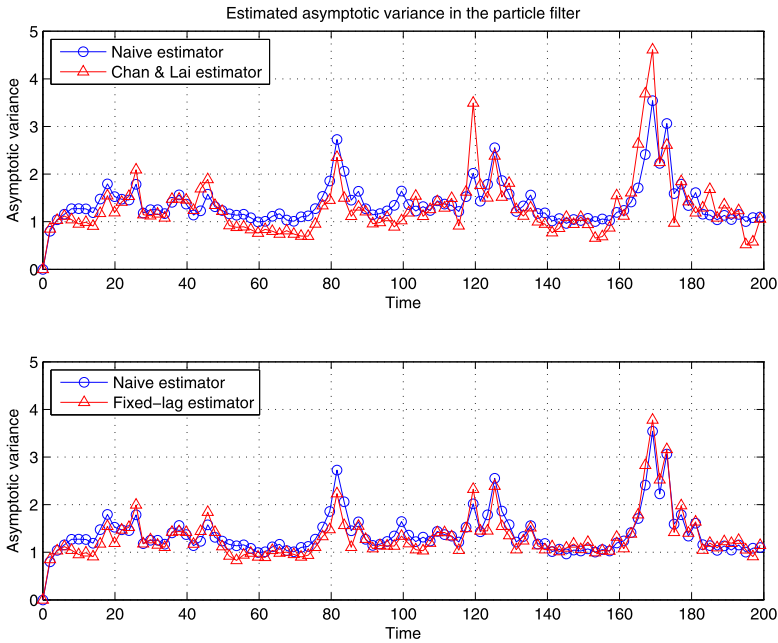
Figure 3. Long-term evolution of estimated asymptotic variances of particle predictor means in the stochastic volatility model (19). For clarity, only every second estimate is plotted. The top panel displays variance estimates (○) produced using the CLE estimator along with reference values obtained as the sample variances (scaled by the number of particles) computed from 1200 independent replicates of the particle predictor mean sequence. The bottom panel displays the analogous plot for estimates (○) produced using the fixed-lag estimator with $\lambda = 20$. The number of particles was set to $N = 5000$ is all cases.

Gaussian models is often used as a testing lab for SMC algorithms. In the setting where $|\varphi| < 1$, the state and observation processes are stationary, with zero mean Gaussian marginal stationary distributions with variances $\tilde{\sigma}^2$ and $\tilde{\sigma}^2 + \sigma_v^2$, respectively, where $\tilde{\sigma}^2 := \sigma_u^2/(1 - \varphi^2)$. We let the former initialise the state process.

Arguing along the lines of Section 5.1, Assumptions (**A1**)–(**A2**) are checked straightforwardly also for this model. We leave the details to the reader.

After having generated, for the parameterisation $(\varphi, \sigma_u, \sigma_v) = (0.98, 0.2, 1)$, an observation record of 600 observations, the experiment in Section 5.1.1 was repeated for the same set of lag sizes λ. As in the stochastic volatility example, the evolution of the $N = 4000$ particles followed the same model dynamics as that generating the observations, and the reference value 1.102 of the variance at $n = 600$ was obtained on the basis of 1000 predictor mean replicates. The outcome is reported in Figure 5 and Table 2, which for this model indicate a somewhat even more robust performance of our estimator with respect to the lag size; indeed, more or less all the lag sizes in the interval $[\![12, 22]\!]$ yield negligible biases, with only a slight increase of variance for the larger ones. According to Table 2, $\lambda = 18$ yields the minimal bias. As in the previous example,
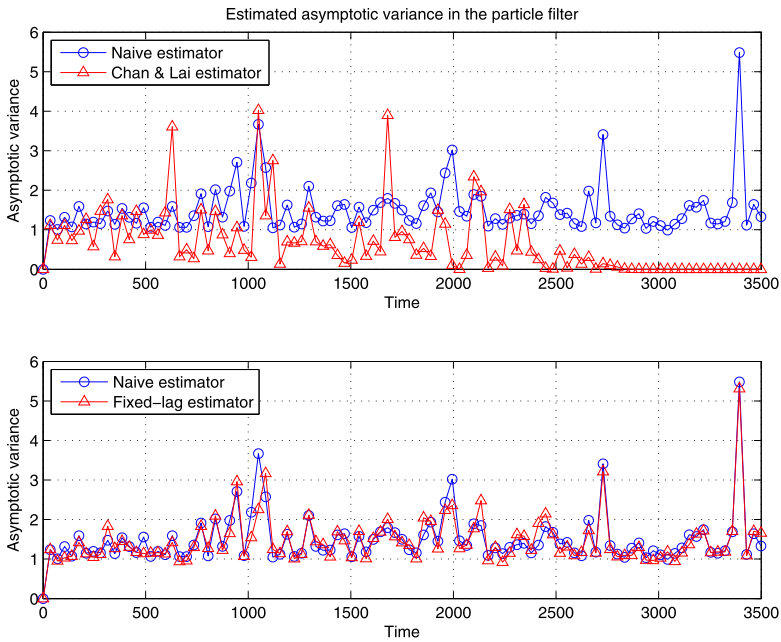
Figure 4. The same plot as in Figure 3, but now for the full observation record comprising 3500 observations. For clarity, only every 35th estimate is plotted.

The CLE (corresponding to the lag size $\lambda = 600$) exhibits very unstable performance due to the relatively high n-to-N ratio, with at least 75% of its estimates falling below the reference value.

Finally, using the lag $\lambda = 18$ extracted from the previous simulation, Algorithm 2 was re-run 150 times on the same observation record $\bar{y}_0^{n-1}$, $n = 600$, each run producing a sequence $\{\eta_N \langle \bar{y}_0^{m-1} \rangle (\mathrm{id})\}_{m=0}^n$ of particle predictor means, a sequence $\{\sigma_{N,m(\lambda)} \langle \bar{y}_0^{m-1} \rangle (\mathrm{id})\}_{m=0}^n$ of fixed-lag variance estimates, and associated approximate 95% confidence intervals $\{I_m\}_{m=0}^n$, each interval given by

$$I_m = \left(\eta_N \langle \bar{y}_0^{m-1} \rangle \mathrm{id} \pm \lambda_{0.025} \frac{\sigma_{N,m(\lambda)} \langle \bar{y}_0^{m-1} \rangle (\mathrm{id})}{\sqrt{N}} \right), \tag{23}$$

where $\lambda_{0.025}$ denotes the 2.5% quantile of the standard Gaussian distribution. As before, N was set to 4000. In the case of effective variance estimation, one may expect each I_m to fail to cover the true predicted mean $\eta \langle \bar{y}_0^{m-1} \rangle (\mathrm{id})$ with a probability close to 5%. Since we are in the setting of a linear Gaussian model, the exact predictor means $\{\eta \langle \bar{y}_0^{m-1} \rangle (\mathrm{id})\}_{m=0}^n$ are accessible through Kalman prediction, and we are thus able to assess the failure rates of the confidence intervals at different time points. Such failure rates are reported in Figure 6, and the red-dashed line indicates the perfect rate 5%. For readability, only every 10th time point is reported. Appealingly, it is clear that the rates fluctuate constantly around 5%, without any notable time dependence. This reconfirms the numeric stability of our fixed-lag variance estimator. The average failure rate across all
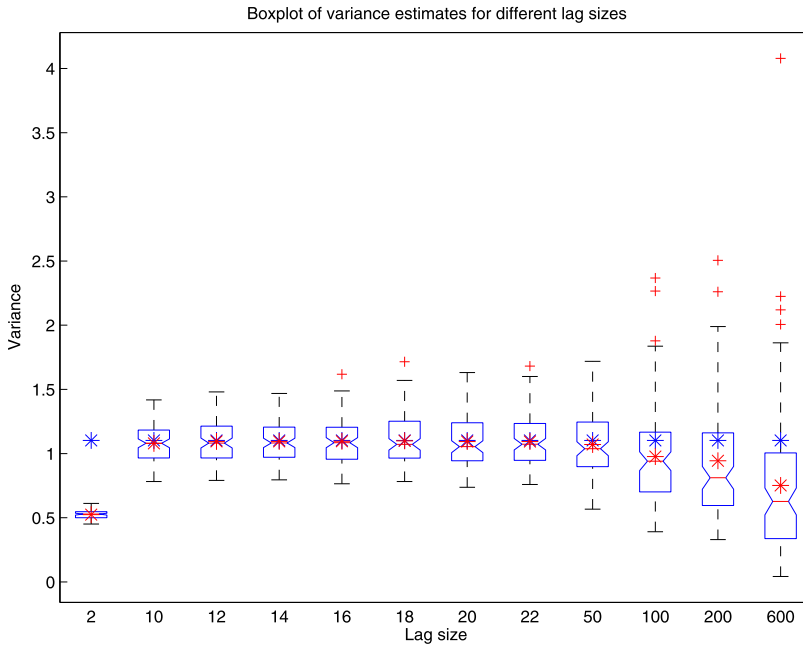
Figure 5. Estimated asymptotic variances of the particle predictor mean at time 600 in the linear Gaussian model (22). The particle population size is $N = 4000$. The boxes are based on 100 replicates of Algorithm 2 and correspond to the different lags $\lambda \in \{2, 10, 12, 14, 16, 18, 20, 22, 50, 100, 200, 600\}$. The reference value 1.102, which is indicated by blue-colored asterisks ($*$), was obtained as the sample variance (scaled by the number of particles) of 1000 independent replicates of the particle predictor mean at time 600 (again, Algorithm 2 used $N = 4000$ particles). Red-colored asterisks indicate average estimates of the boxes.

600 time points was 5.5%. The fact that the average failure rate is slightly above the perfect rate 5% is in line with the fact that the bias of our estimator is always positive, as underestimation of variance leads to more narrow confidence bounds and, consequently, higher failure rates. One way of hedging against underestimation of variance could be to replace Gaussian quantiles by the quantiles of some Student's t-distribution with a moderate number of degrees of freedom.

6. Conclusion

The estimator of the SMC asymptotic variance that we propose is a natural modification of the CLE introduced in [2]. As in [17,21], the main idea is to reduce the degree of genealogical tracing, which has a devastating effect on the CLE's numerical stability, at the cost of a small bias, which may be controlled using the forgetting properties of the model. That this measure stabilises numerically the estimator in the long run is confirmed by our theoretical results in Section 4, which are obtained under—what we believe—minimal model assumptions being satisfied also

Table 2. Means and standard deviations of the variance estimates reported in Figure 5. The reference value is 1.102

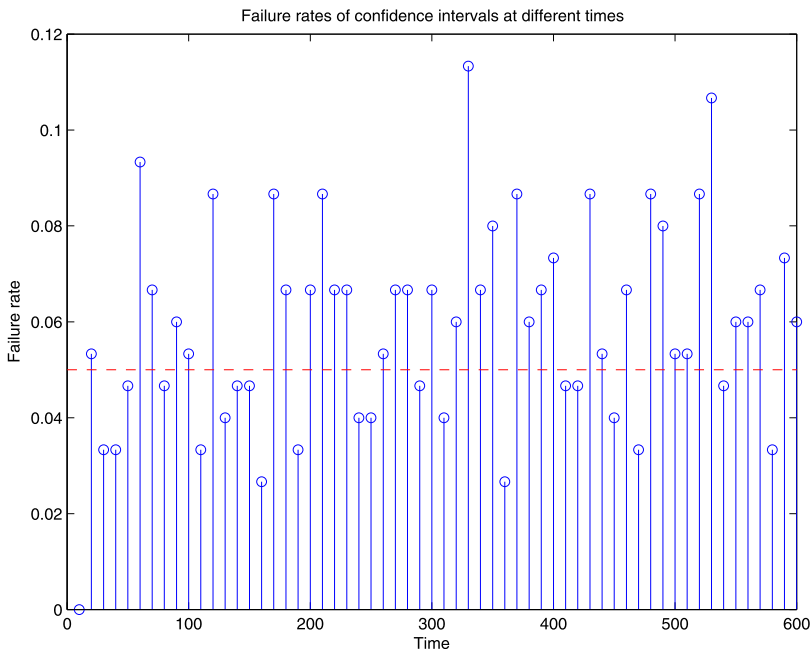| λ	Mean	St. dev.
2	0.524	0.035
10	1.080	0.157
12	1.095	0.163
14	1.095	0.162
16	1.096	0.175
18	1.099	0.190
20	1.094	0.198
22	1.093	0.202
50	1.071	0.246
100	0.976	0.370
200	0.944	0.471
600	0.751	0.593

Figure 6. Failure rates of confidence bounds (23) at time points $m \in \{10, 20, 30, \ldots, 600\}$. The red-dashed line indicates the perfect rate 5%. The failure rate estimates are based on 150 runs of Algorithm 2 with $N = 4000$ particles.

for many models with possibly non-compact state space. The fact that the bias can be shown to decrease geometrically fast as the lag increases indicates that tight control of the bias is possible also for moderately large particle sample sizes. This is approved by our numerical experiments in Section 5, which report, in the examples under consideration, a negligible bias already for some thousands of particles.

Needless to say, the success of our approach depends highly on the interplay between the forgetting properties of the model, the particle sample size, and the choice of the lag. Adaptive lag design is hence a natural direction for future research. Moreover, as our estimator provides numerically stable estimates of the asymptotic variance, it should be highly useful for online SMC sample size adaptation. Here one natural approach could be to estimate the variance of the next time step using a part of the particle population (*pilot sampling*) and then "refuel" the particle system at time steps of high variance (here the techniques developed in [19], where the authors consider the SMC sample allocation problem in the batch mode, should be useful for the theoretical analysis). Finally, casting, using the results obtained in [20], our estimator and analysis into the framework of *Rao–Blackwellised SMC algorithms*, should be of high relevance for high-dimensional applications.

Appendix: Proofs

A.1. Proof of Proposition 5

The proof of Proposition 5 relies on the machinery developed in [19], from which we adopt the following definitions. Throughout this section, let $n \in \mathbb{N}$ and $\bar{z}_0^{n-1} \in \mathsf{Z}^n$ be picked arbitrarily.

- Denote by $\mathsf{B}_n := \{0, 1\}^{n+1}$ the space of binary strings of length $n + 1$. The *zero string* of length $n + 1$ is denoted by 0_n and for $m \in [\![0, n]\!]$, 1_m denotes a *unit string* of length $n + 1$ with 1 on position m (with positions indexed from 0) and zeros everywhere else.
- For a given string $\bar{b}_0^n \in \mathsf{B}_n$, a Markov chain $\{(X_m, X_m')\}_{m=0}^n$ on $(\mathsf{X}^2, \mathcal{X}^{\otimes 2})$ is defined as follows. If $b_0 = 0$, then $(X_0, X_0') \sim \chi^{\otimes 2}$; otherwise, if $b_0 = 1$, $X_0' = X_0 \sim \chi$ (the initial distribution). After this, if $b_{m+1} = 0$, $X_{m+1} \sim \mathbf{M}(X_m, \cdot)$ and $X_{m+1}' \sim \mathbf{M}(X_m', \cdot)$ conditionally independently; otherwise, if $b_{m+1} = 1$, $X_{m+1}' = X_{m+1} \sim \mathbf{M}(X_m, \cdot)$.
- With $\mathbb{E}_{\bar{b}_0^n}$ denoting the expectation under the law of $\{(X_m, X_m')\}_{m=0}^n$, we define, for all $\bar{b}_0^n \in \mathsf{B}_n$, the measures

$$\mu_{\bar{b}_0^n}\langle \bar{z}_0^{n-1} \rangle : \mathcal{X}^{\otimes 2} \ni A \mapsto \mathbb{E}_{\bar{b}_0^n}\left[\mathbb{1}_A(X_n, X_n') \prod_{m=0}^{n-1} g\langle z_m \rangle(X_m) g\langle z_m \rangle(X_m') \right].$$

Note that for all $h \in \mathbb{F}(\mathcal{X})$ it holds that $\mu_{0_n}\langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} = (\chi \mathbf{Q}\langle \bar{z}_0^{n-1} \rangle h)^2$ and $\mu_{1_m}\langle \bar{z}_0^{n-1} \rangle \times h^{\otimes 2} = \chi \mathbf{Q}\langle \bar{z}_0^{n-1} \rangle \mathbb{1}_{\mathsf{X}} \times \chi \mathbf{Q}\langle \bar{z}_0^{n-1} \rangle (\mathbf{Q}\langle \bar{z}_m^{n-1} \rangle h)^2$, and defining

$$\upsilon_{m,n}\langle \bar{z}_0^{n-1} \rangle : \mathbb{F}(\mathcal{X}) \ni h \mapsto \frac{\mu_{1_m}\langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} - \mu_{0_n}\langle \bar{z}_0^{n-1} \rangle h^{\otimes 2}}{(\chi \mathbf{Q}\langle \bar{z}_0^{n-1} \rangle \mathbb{1}_{\mathsf{X}})^2}$$

yields for all $h \in \mathbb{F}(\mathcal{X})$,

$$v_{m,n}\langle \bar{z}_0^{n-1}\rangle(h) = \frac{\eta\langle\bar{z}_0^{m-1}\rangle(\mathbf{Q}\langle\bar{z}_m^{n-1}\rangle h)^2}{(\eta\langle\bar{z}_0^{m-1}\rangle\mathbf{Q}\langle\bar{z}_m^{n-1}\rangle\mathbb{1}_\mathsf{X})^2} - \left(\eta\langle\bar{z}_0^{n-1}\rangle h\right)^2$$

and, consequently, for all $\ell \in [\![0, n]\!]$,

$$\sigma_\ell^2\langle\bar{z}_0^{n-1}\rangle(h) = \sum_{m=\ell}^{n} v_{m,n}\langle\bar{z}_0^{n-1}\rangle\left(h - \eta\langle\bar{z}_0^{n-1}\rangle h\right). \tag{24}$$

- For all $N \in \mathbb{N}^*$, let $\mathcal{F}_n := \sigma(\{\xi_0^i\}_{i=1}^N, \{\xi_m^i, I_m^i\}_{i=1}^N; m \in [\![1, n]\!])$ be the σ-field generated by the output of Algorithm 1 during the first n iterations. Conditionally on $\mathcal{F}_n$, a genealogical trace $\bar{G}_0^n$ is formed backwards in time by, first, drawing G_n uniformly over $[\![1, N]\!]$ and, second, setting $G_m = I_{m+1}^{G_{m+1}}$ for all $m \in [\![0, n-1]\!]$. In addition, a parallel trace $\bar{G}'_0^n$ is formed by, first, drawing G'_n uniformly over $[\![1, N]\!]$ and, second, letting $G'_m = I_{m+1}^{G'_{m+1}}$ if $G'_{m+1} \neq G_{m+1}$ or $G'_m \sim \mathsf{Cat}(\{\omega_m^i/\Omega_m\}_{i=1}^N)$ otherwise.

The proof of the following lemma follows closely that of [19], Lemma 4 and is hence omitted. Define for all $\ell \in [\![0, n]\!]$ and $\bar{b}_0^\ell \in \mathsf{B}_\ell$,

$$\mathsf{I}_\ell(\bar{b}_0^\ell) := \left\{(\bar{k}_0^\ell, \bar{k}'_0^\ell) \in [\![1, N]\!]^{2(\ell+1)} : \text{for all } \ell' \in [\![0, \ell]\!], k_{\ell'} = k'_{\ell'} \Leftrightarrow b_{\ell'} = 1\right\}.$$

Lemma 12. *For all $N \in \mathbb{N}^*$ and $m \in [\![0, n]\!]$,*

$$\left\{E_{m,n}^{G_n} \neq E_{m,n}^{G'_n}\right\} = \left\{(\bar{G}_m^n, \bar{G}'_m^n) \in \mathsf{I}_{n-m}(0_{n-m})\right\}.$$

In addition, define, for all $N \in \mathbb{N}^*$, the measures

$$\gamma_N\langle\bar{z}_0^{n-1}\rangle : \mathcal{X} \ni A \mapsto \frac{1}{N^{n+1}}\left(\prod_{m=0}^{n-1}\Omega_m\right)\sum_{i=1}^N \mathbb{1}_A(\xi_n^i) \tag{25}$$

and

$$\mu_{N,\bar{b}_0^n}\langle\bar{z}_0^{n-1}\rangle : \mathcal{X}^{\otimes 2} \ni A \mapsto N^{\#_1(\bar{b}_0^n)}\left(\frac{N}{N-1}\right)^{\#_0(\bar{b}_0^n)}\left(\gamma_N\langle\bar{z}_0^{n-1}\rangle\mathbb{1}_\mathsf{X}\right)^2$$

$$\times \mathbb{E}\left[\mathbb{1}_A(\xi_n^{G_n}, \xi_n^{G'_n})\mathbb{1}\left\{(\bar{G}_0^n, \bar{G}'_0^n) \in \mathsf{I}_n(\bar{b}_0^n)\right\} \mid \mathcal{F}_n\right], \tag{26}$$

where $\#_1(\bar{b}_0^n) := \sum_{m=0}^n b_m$ and $\#_0(\bar{b}_0^n) := n + 1 - \#_1(\bar{b}_0^n)$ denote the numbers of ones and zeros in $\bar{b}_0^n$, respectively. Note that (25) implies that for all $h \in \mathbb{F}(\mathcal{X})$, $\eta_N\langle\bar{z}_0^{n-1}\rangle h = \gamma_N\langle\bar{z}_0^{n-1}\rangle h/\gamma_N\langle\bar{z}_0^{n-1}\rangle\mathbb{1}_\mathsf{X}$.

The following lemma, where first part is established in [19], Theorem 2 and the last part is a standard result (see, e.g., [10] for results on the weak consistency of SMC), will be instrumental.

Lemma 13. *For all $\bar{b}_0^n \in \mathsf{B}_n$ and $h \in \mathbb{F}(\mathcal{X}^{\otimes 2})$, as $N \to \infty$,*

$$\mu_{N,\bar{b}_0^n}\langle \bar{z}_0^{n-1}\rangle h \xrightarrow{\mathbb{P}} \mu_{\bar{b}_0^n}\langle \bar{z}_0^{n-1}\rangle h.$$

In addition, for all $h \in \mathbb{F}(\mathcal{X})$,

$$\gamma_N\langle \bar{z}_0^{n-1}\rangle h \xrightarrow{\mathbb{P}} \chi \mathbf{Q}\langle \bar{z}_0^{n-1}\rangle h.$$

Proof of Proposition 5. Fix $\ell \in [\![0, n]\!]$ and define for all $N \in \mathbb{N}^*$,

$$\varphi_{N,\ell}\langle \bar{z}_0^{n-1}\rangle : \mathcal{X}^{\otimes 2} \ni A \mapsto \left(\gamma_N\langle \bar{z}_0^{n-1}\rangle \mathbb{1}_{\mathsf{X}}\right)^2$$
$$\times \mathbb{E}\left[\mathbb{1}_A\left(\xi_n^{G_n}, \xi_n^{G'_n}\right)\mathbb{1}\left\{\left(\bar{G}_\ell^n, \bar{G}'^n_\ell\right) \in \mathsf{I}_{n-\ell}(0_{n-\ell})\right\} \mid \mathcal{F}_n\right].$$

First, note that by Lemma 12, since G_n and G'_n are conditionally independent and uniformly distributed over $[\![1, N]\!]$, for all $h \in \mathbb{F}(\mathcal{X})$,

$$\frac{1}{N^2}\sum_{i=1}^N\left(\sum_{j:E_{\ell,n}^j=i} h\left(\xi_n^j\right)\right)^2 = \frac{1}{N^2}\sum_{(i,j):E_{\ell,n}^i=E_{\ell,n}^j} h\left(\xi_n^i\right)h\left(\xi_n^j\right)$$

$$= \left(\eta_N\langle \bar{z}_0^{n-1}\rangle h\right)^2 - \frac{1}{N^2}\sum_{(i,j):E_{\ell,n}^i\neq E_{\ell,n}^j} h\left(\xi_n^i\right)h\left(\xi_n^j\right) \qquad (27)$$

$$= \left(\eta_N\langle \bar{z}_0^{n-1}\rangle h\right)^2 - \frac{\varphi_{N,\ell}\langle \bar{z}_0^{n-1}\rangle h^{\otimes 2}}{\left(\gamma_N\langle \bar{z}_0^{n-1}\rangle \mathbb{1}_{\mathsf{X}}\right)^2}.$$

It is hence enough to prove that for all $N \in \mathbb{N}^*$ and $h \in \mathbb{F}(\mathcal{X})$,

$$N\left\{\left(\gamma_N\langle \bar{z}_0^{n-1}\rangle h\right)^2 - \varphi_{N,\ell}\langle \bar{z}_0^{n-1}\rangle h^{\otimes 2}\right\}$$
$$= \sum_{m=\ell}^n \left(\mu_{N,1_m}\langle \bar{z}_0^{n-1}\rangle h^{\otimes 2} - \mu_{N,0_n}\langle \bar{z}_0^{n-1}\rangle h^{\otimes 2}\right) + (n - \ell + 1)\left(\gamma_N\langle \bar{z}_0^{n-1}\rangle h\right)^2 \qquad (28)$$
$$+ \|h\|_\infty^2 \mathcal{O}(N^{-1}),$$

where the $\mathcal{O}(N^{-2})$ term does not depend on h; indeed, along the lines of the proof of [19], Theorem 1, Lemma 13 implies that for all $\bar{b}_0^n \in \mathsf{B}_n$, as $N \to \infty$,

$$\mu_{N,\bar{b}_0^n}\langle \bar{z}_0^{n-1}\rangle\left\{h - \eta_N\langle \bar{z}_0^{n-1}\rangle h\right\}^{\otimes 2} \xrightarrow{\mathbb{P}} \mu_{\bar{b}_0^n}\langle \bar{z}_0^{n-1}\rangle\left\{h - \eta\langle \bar{z}_0^{n-1}\rangle h\right\}^{\otimes 2},$$

and (28) hence yields, with $\ell = n(\lambda)$ and when combined with (27) and (24), again as $N \to \infty$,

$$
\sigma_{N,\lambda}^2 \langle \bar{z}_0^{n-1} \rangle (h)
$$

$$
= \sum_{m=n(\lambda)}^n \frac{\mu_{N,1_m} \langle \bar{z}_0^{n-1} \rangle \{h - \eta_N \langle \bar{z}_0^{n-1} \rangle h\}^{\otimes 2} - \mu_{N,0_n} \langle \bar{z}_0^{n-1} \rangle \{h - \eta_N \langle \bar{z}_0^{n-1} \rangle h\}^{\otimes 2}}{(\gamma_N \langle \bar{z}_0^{n-1} \rangle \mathbb{1}_X)^2}
\tag{29}
$$

$$
+ \|h\|_\infty^2 \mathcal{O}(N^{-1}) \xrightarrow{\mathbb{P}} \sigma_{n(\lambda)}^2 \langle \bar{z}_0^{n-1} \rangle (h).
$$

In order to establish (28), write, using that G_n and G'_n are, given $\mathcal{F}_n$, conditionally independent and uniformly distributed over $[\![1, N]\!]$,

$$
\left(\gamma_N \langle \bar{z}_0^{n-1} \rangle h \right)^2 - \varphi_{N,\ell} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2}
$$

$$
= \left(\gamma_N \langle \bar{z}_0^{n-1} \rangle \mathbb{1}_X \right)^2 \sum_{\bar{b}_0^n \in \mathsf{B}_n} \mathbb{E}\left[h\left(\xi_n^{G_n}\right) h\left(\xi_n^{G'_n}\right) \mathbb{1}\left\{ \left(\bar{G}_0^n, \bar{G}'^n_0\right) \in \mathsf{I}_n\left(\bar{b}_0^n\right) \right\} \mid \mathcal{F}_n \right]
\tag{30}
$$

$$
- \left(\gamma_N \langle \bar{z}_0^{n-1} \rangle \mathbb{1}_X \right)^2 \sum_{\bar{b}_0^n \in \mathsf{B}_n : \bar{b}_\ell^n = 0_{n-\ell}} \mathbb{E}\left[h\left(\xi_n^{G_n}\right) h\left(\xi_n^{G'_n}\right) \mathbb{1}\left\{ \left(\bar{G}_0^n, \bar{G}'^n_0\right) \in \mathsf{I}_n\left(\bar{b}_0^n\right) \right\} \mid \mathcal{F}_n \right]
$$

and note that, by definition (26),

$$
\left(\gamma_N \langle \bar{z}_0^{n-1} \rangle \mathbb{1}_X \right)^2 \sum_{\bar{b}_0^n \in \mathsf{B}_n} \mathbb{E}\left[h\left(\xi_n^{G_n}\right) h\left(\xi_n^{G'_n}\right) \mathbb{1}\left\{ \left(\bar{G}_0^n, \bar{G}'^n_0\right) \in \mathsf{I}_n\left(\bar{b}_0^n\right) \right\} \mid \mathcal{F}_n \right]
$$

$$
= \frac{1}{N} \sum_{m=0}^n \mu_{N,1_m} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} + \left(1 - \frac{1}{N}\right)^{n+1} \mu_{N,0_n} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} + \|h\|_\infty^2 \mathcal{O}(N^{-2})
\tag{31}
$$

$$
= \frac{1}{N} \sum_{m=0}^n \left(\mu_{N,1_m} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} - \mu_{N,0_n} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} \right) + \mu_{N,0_n} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} + \|h\|_\infty^2 \mathcal{O}(N^{-2}).
$$

Similarly,

$$
\left(\gamma_N \langle \bar{z}_0^{n-1} \rangle \mathbb{1}_X \right)^2 \sum_{\bar{b}_0^n \in \mathsf{B}_n : \bar{b}_\ell^n = 0_{n-\ell}} \mathbb{E}\left[h\left(\xi_n^{G_n}\right) h\left(\xi_n^{G'_n}\right) \mathbb{1}\left\{ \left(\bar{G}_0^n, \bar{G}'^n_0\right) \in \mathsf{I}_n\left(\bar{b}_0^n\right) \right\} \mid \mathcal{F}_n \right]
$$

$$
= \frac{1}{N} \sum_{m=0}^{\ell-1} \left(\mu_{N,1_m} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} - \mu_{N,0_n} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} \right) + \mu_{N,0_n} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2}
\tag{32}
$$

$$
- \frac{n - \ell + 1}{N} \mu_{N,0_n} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} + \|h\|_\infty^2 \mathcal{O}(N^{-2}),
$$

and combining (30), (31), (32), and the fact that $\mu_{0_n} \langle \bar{z}_0^{n-1} \rangle h^{\otimes 2} = (\chi \mathbf{Q} \langle \bar{z}_0^{n-1} \rangle h)^2$ yields (28). This completes the proof. $\qquad\square$

A.2. Proof of Theorem 6

In the proof of Theorem 6, the asymptotic bias is bounded using the time-shift approach taken in [12], Theorem 10. Even though the theoretical analysis provided in [12] is cast into the framework of general state-space models, it never makes use of the fact that g is a normalised transition density. As stressed in [12], Remark 1, it is hence directly applicable to the framework of randomly perturbed Feynman–Kac models in Section 2.2.

Proof. Pick arbitrarily $n \in \mathbb{N}$, $\lambda \in \mathbb{N}$, $h \in \mathbb{F}(\mathcal{X})$, and $\chi \in \mathbb{M}(D, r)$. By defining, for all $m \in \mathbb{N}$, $\ell \in [\![0, m-1]\!]$, $\bar{z}_\ell^{m-1} \in \mathsf{Z}^{m-\ell}$, and measures $(\mu, \mu') \in \mathbb{M}(\mathcal{X})^2$,

$$
\Delta_{\mu, \mu'} \langle \bar{z}_\ell^{m-1} \rangle : \mathbb{F}(\mathcal{X})^2 \ni (h, h') \mapsto \mu \mathbf{Q} \langle \bar{z}_\ell^{m-1} \rangle h \times \mu' \mathbf{Q} \langle \bar{z}_\ell^{m-1} \rangle h' \\
- \mu \mathbf{Q} \langle \bar{z}_\ell^{m-1} \rangle h' \times \mu' \mathbf{Q} \langle \bar{z}_\ell^{m-1} \rangle h,
\tag{33}
$$

we may, using the identity

$$
\eta \langle \bar{\zeta}_0^{m-1} \rangle \mathbf{Q} \langle \bar{\zeta}_m^{n-1} \rangle \mathbb{1}_{\mathsf{X}} = \frac{\chi \mathbf{Q} \langle \bar{\zeta}_0^{n-1} \rangle \mathbb{1}_{\mathsf{X}}}{\chi \mathbf{Q} \langle \bar{\zeta}_0^{m-1} \rangle \mathbb{1}_{\mathsf{X}}} = \prod_{\ell=m}^{n-1} \frac{\chi \mathbf{Q} \langle \bar{\zeta}_0^{\ell} \rangle \mathbb{1}_{\mathsf{X}}}{\chi \mathbf{Q} \langle \bar{\zeta}_0^{\ell-1} \rangle \mathbb{1}_{\mathsf{X}}} = \prod_{\ell=m}^{n-1} \eta \langle \bar{\zeta}_0^{\ell-1} \rangle g \langle \zeta_\ell \rangle,
$$

for all $m \in [\![0, n]\!]$, write the asymptotic bias at time n as

$$
\beta_\lambda \langle \bar{\zeta}_0^{n-1} \rangle (h) = \sum_{m=0}^{n(\lambda)-1} \int \eta \langle \bar{\zeta}_0^{m-1} \rangle (\mathrm{d}x) \left(\frac{\Delta_{\delta_x, \eta \langle \bar{\zeta}_0^{m-1} \rangle} \langle \bar{\zeta}_m^{n-1} \rangle (h, \mathbb{1}_{\mathsf{X}})}{[\prod_{\ell=m}^{n-1} \eta \langle \bar{\zeta}_0^{\ell-1} \rangle g \langle \zeta_\ell \rangle]^2} \right)^2.
\tag{34}
$$

In addition, under the assumptions of the theorem, [11], Proposition 1 provides the existence of a function $\pi : \mathsf{Z}^\infty \to \mathbb{R}$ such that for all initial distributions $\chi \in \mathbb{M}(D, r)$,

$$
\lim_{m \to \infty} \eta \langle \bar{\zeta}_{-m}^{-1} \rangle g \langle \zeta_0 \rangle = \pi \langle \bar{\zeta}_{-\infty}^0 \rangle, \qquad \mathbb{P}\text{-a.s.}
$$

Since the perturbations $\{\zeta_n\}_{n \in \mathbb{Z}}$ are stationary, the distribution of $\beta_\lambda \langle \bar{\zeta}_0^{n-1} \rangle (h)$ coincides with that of the time-shifted bias $\beta_\lambda \langle \bar{\zeta}_{-n}^{-1} \rangle (h)$, and a key step in the present proof is to express, via (34), the latter as

$$
\beta_\lambda \langle \bar{\zeta}_{-n}^{-1} \rangle (h)
$$
$$
= \sum_{m=0}^{n(\lambda)-1} \int \eta \langle \bar{\zeta}_{-n}^{-n+n(\lambda)-m-2} \rangle (\mathrm{d}x) \left(\frac{\Delta_{\delta_x, \eta \langle \bar{\zeta}_{-n}^{-n+n(\lambda)-m-2} \rangle} \langle \bar{\zeta}_{-n+n(\lambda)-m-1}^{-1} \rangle (h, \mathbb{1}_{\mathsf{X}})}{[\prod_{\ell=1}^{n-n(\lambda)+m+1} \eta \langle \bar{\zeta}_{-n}^{-\ell-1} \rangle g \langle \zeta_{-\ell} \rangle]^2} \right)^2.
$$

We hence obtain the bound

$$
\beta_\lambda \langle \bar{\zeta}_{-n}^{-1} \rangle (h) \le A_n \times B_n,
\tag{35}
$$

where

$$A_n := \left(\sup_{(k,m)\in\mathbb{Z}^2:\,-n\le k\le m} \prod_{\ell=k}^{m-1} \frac{\pi\langle \bar{\zeta}_{-\infty}^{-\ell}\rangle}{\eta\langle \bar{\zeta}_{-n}^{-\ell-1}\rangle g\langle \zeta_\ell\rangle} \right)^4,$$

$$B_n := \sum_{m=0}^{n(\lambda)-1} \left(\frac{\sup_{x\in\mathsf{X}} |\Delta_{\delta_x,\eta\langle \bar{\zeta}_{-n}^{-n+n(\lambda)-m-2}\rangle}\langle \bar{\zeta}_{-n+n(\lambda)-m-1}^{-1}\rangle(h,\mathbb{1}_\mathsf{X})|}{[\prod_{\ell=1}^{n-n(\lambda)+m+1}\pi\langle \bar{\zeta}_{-\infty}^{-\ell}\rangle]^2} \right)^2.$$

To bound uniformly the sequence $\{B_n\}_{n\in\mathbb{N}}$, decompose each term according to

$$\frac{\sup_{x\in\mathsf{X}} |\Delta_{\delta_x,\eta\langle \bar{\zeta}_{-n}^{-n+n(\lambda)-m-2}\rangle}\langle \bar{\zeta}_{-n+n(\lambda)-m-1}^{-1}\rangle(h,\mathbb{1}_\mathsf{X})|}{[\prod_{\ell=1}^{n-n(\lambda)+m+1}\pi\langle \bar{\zeta}_{-\infty}^{-\ell}\rangle]^2}$$

$$= \left(\frac{\|\mathbf{Q}\langle \bar{\zeta}_{-n+n(\lambda)-m-1}^{-1}\rangle\mathbb{1}_\mathsf{X}\|_\infty}{\prod_{\ell=1}^{n-n(\lambda)+m+1}\pi\langle \bar{\zeta}_{-\infty}^{-\ell}\rangle} \right)^2 \tag{36}$$

$$\times \frac{\sup_{x\in\mathsf{X}} |\Delta_{\delta_x,\eta\langle \bar{\zeta}_{-n}^{-n+n(\lambda)-m-2}\rangle}\langle \bar{\zeta}_{-n+n(\lambda)-m-1}^{-1}\rangle(h,\mathbb{1}_\mathsf{X})|}{\|\mathbf{Q}\langle \bar{\zeta}_{-n+n(\lambda)-m-1}^{-1}\rangle\mathbb{1}_\mathsf{X}\|_\infty^2}.$$

We consider separately the two factors of (36). First,

$$\left(\frac{\|\mathbf{Q}\langle \bar{\zeta}_{-n+n(\lambda)-m-1}^{-1}\rangle\mathbb{1}_\mathsf{X}\|_\infty}{\prod_{\ell=1}^{n-n(\lambda)+m+1}\pi\langle \bar{\zeta}_{-\infty}^{-\ell}\rangle} \right)^2 = \exp\{(n-n(\lambda)+m+1)\varepsilon_{n-n(\lambda)+m+1}\}, \tag{37}$$

with

$$\varepsilon_k := \frac{2}{k}\left(\ln\|\mathbf{Q}\langle \bar{\zeta}_{-k}^{-1}\rangle\mathbb{1}_\mathsf{X}\|_\infty - \sum_{\ell=1}^k \ln\pi\langle \bar{\zeta}_{-\infty}^{-\ell}\rangle \right)$$

being independent of χ for all $k\in\mathbb{N}^*$. By [12], Lemma 17, $\varepsilon_k \to 0$, $\mathbb{P}$-a.s., as $k\to\infty$, which implies that (37) grows at most subgeometrically fast with m. In addition, by [12], Proposition 16(iii) there exists a constant $\rho\in(0,1)$ and a $\mathbb{P}$-a.s. finite random variable D such that for all n and m, all $h\in\mathbb{F}(\mathcal{X})$, and all $\chi\in\mathbb{M}(\mathcal{X})$, $\mathbb{P}$-a.s.,

$$\frac{\sup_{x\in\mathsf{X}} |\Delta_{\delta_x,\eta\langle \bar{\zeta}_{-n}^{-n+n(\lambda)-m-2}\rangle}\langle \bar{\zeta}_{-n+n(\lambda)-m-1}^{-1}\rangle(h,\mathbb{1}_\mathsf{X})|}{\|\mathbf{Q}\langle \bar{\zeta}_{-n+n(\lambda)-m-1}^{-1}\rangle\mathbb{1}_\mathsf{X}\|_\infty^2} \le D\rho^{n-n(\lambda)+m+1}\|h\|_\infty.$$

Thus, $\mathbb{P}$-a.s.,

$$B_n \le D^2\|h\|_\infty^2 \sum_{m=0}^{n(\lambda)-1} \rho^{2(n-n(\lambda)+m+1)} \exp\{2(n-n(\lambda)+m+1)\varepsilon_{n-n(\lambda)+m+1}\}.$$

If $n \leq \lambda$, then $n(\lambda) = 0$, and the bias vanishes. Thus, we assume in the following that $\lambda < n$, which means that $n(\lambda) = n - \lambda$. Then, by the Cauchy–Schwarz inequality, $\mathbb{P}$-a.s.,

$$
\begin{aligned}
B_n &\leq D^2 \|h\|_\infty^2 \sum_{m=\lambda+1}^{\infty} \rho^{2m} \exp(2m\varepsilon_m) \\
&\leq D^2 \|h\|_\infty^2 \left(\sum_{m=\lambda+1}^{\infty} \rho^{2m} \right)^{1/2} \left(\sum_{m=\lambda+1}^{\infty} \rho^{2m} \exp(4m\varepsilon_m) \right)^{1/2} \\
&\leq D^2 \|h\|_\infty^2 \rho^{\lambda+1} \left(\sum_{m=0}^{\infty} \rho^{2m} \right)^{1/2} \left(\sum_{m=0}^{\infty} \rho^{2m} \exp(4m\varepsilon_m) \right)^{1/2} \\
&= C \|h\|_\infty^2 \rho^{\lambda+1},
\end{aligned}
\tag{38}
$$

where random variable

$$
C := D^2 \left(\sum_{m=0}^{\infty} \rho^{2m} \right)^{1/2} \left(\sum_{m=0}^{\infty} \rho^{2m} \exp(4m\varepsilon_m) \right)^{1/2}
$$

is $\mathbb{P}$-a.s. finite and independent of λ, h, and χ. For $c \in \mathbb{R}_+$, write, using (35) and (38),

$$
\mathbb{P}\left(\frac{\beta_\lambda \langle \bar{\zeta}_0^{n-1} \rangle (h)}{\rho^{\lambda+1} \|h\|_\infty^2} > c \right) = \mathbb{P}\left(\frac{\beta_\lambda \langle \bar{\zeta}_{-n}^{-1} \rangle (h)}{\rho^{\lambda+1} \|h\|_\infty^2} > c \right) \leq \mathbb{P}(A_n C > c),
$$

where the probability on the right hand side is again independent of λ, h, or χ. Thus, using the stationarity of $\{A_n\}_{n \in \mathbb{N}}$,

$$
\mathbb{P}\left(\frac{\beta_\lambda \langle \bar{\zeta}_0^{n-1} \rangle (h)}{\rho^{\lambda+1} \|h\|_\infty^2} > c \right) \leq \mathbb{P}(A_0 > c^{1/2}) + \mathbb{P}(C > c^{1/2}),
$$

where the right-hand side does not depend on n. Now, the $\mathbb{P}$-a.s. finiteness of A_0 was established as a part of the proof of [12], Theorem 10. Consequently, as also C is $\mathbb{P}$-a.s. finite, there exists, for all $k \in \mathbb{N}^*$, a constant $c_k \in \mathbb{R}_+$, independent of λ, h, and χ, such that the probabilities $\mathbb{P}(A_0 > c_k^{1/2})$ and $\mathbb{P}(C > c_k^{1/2})$ are both bounded by $1/(2k)$. This completes the proof. $\qquad \square$

A.3. Proof of Proposition 8

Proof. The proof consists mainly in combining some of the equalities in the proof of Proposition 5 with the identity

$$
\begin{aligned}
\{ g \langle z_n \rangle (h - \bar{\eta}_N \langle \bar{z}_0^n \rangle h) \}^{\otimes 2} &= (g \langle z_n \rangle h)^{\otimes 2} - \{ (g \langle z_n \rangle h) \otimes g \langle z_n \rangle \} \bar{\eta}_N \langle \bar{z}_0^n \rangle h \\
&\quad - \{ g \langle z_n \rangle \otimes (g \langle z_n \rangle h) \} \bar{\eta}_N \langle \bar{z}_0^n \rangle h + g \langle z_n \rangle^{\otimes 2} (\bar{\eta}_N \langle \bar{z}_0^n \rangle h)^2.
\end{aligned}
\tag{39}
$$

More specifically, as it holds that

$$\eta_N \langle \bar{z}_0^{n-1} \rangle \big(g \langle z_n \rangle \{ h - \bar{\eta}_N \langle \bar{z}_0^n \rangle h \} \big) = 0, \tag{40}$$

by reusing the equality in (29),

$$\begin{aligned}
\sigma_{N,\lambda}^2 &\langle \bar{z}_0^{n-1} \rangle \big(g \langle z_n \rangle \{ h - \bar{\eta}_N \langle \bar{z}_0^n \rangle h \} \big) \\
&= \sum_{m=n(\lambda)}^{n} \frac{\mu_{N,1_m} \langle \bar{z}_0^{n-1} \rangle \{ g \langle z_n \rangle (h - \bar{\eta}_N \langle \bar{z}_0^n \rangle h) \}^{\otimes 2} - \mu_{N,0_n} \langle \bar{z}_0^{n-1} \rangle \{ g \langle z_n \rangle (h - \bar{\eta}_N \langle \bar{z}_0^n \rangle h) \}^{\otimes 2}}{(\gamma_N \langle \bar{z}_0^{n-1} \rangle \mathbb{1}_{\mathsf{X}})^2} \\
&\quad + \big\| g \langle z_n \rangle \big\|_{\infty}^2 \| h \|_{\infty}^2 \mathcal{O}(N^{-1}).
\end{aligned} \tag{41}$$

Now, write, using (39), for all $\bar{b}_0^n \in \mathsf{B}_n$,

$$\begin{aligned}
\mu_{N,\bar{b}_0^n} &\langle \bar{z}_0^{n-1} \rangle \big\{ g \langle z_n \rangle \big(h - \bar{\eta}_N \langle \bar{z}_0^n \rangle h \big) \big\}^{\otimes 2} \\
&= \mu_{N,\bar{b}_0^n} \langle \bar{z}_0^{n-1} \rangle \big(g \langle z_n \rangle h \big)^{\otimes 2} \\
&\quad - \mu_{N,\bar{b}_0^n} \langle \bar{z}_0^{n-1} \rangle \big\{ \big(g \langle z_n \rangle h \big) \otimes g \langle z_n \rangle \big\} \bar{\eta}_N \langle \bar{z}_0^n \rangle h - \mu_{N,\bar{b}_0^n} \langle \bar{z}_0^{n-1} \rangle \big\{ g \langle z_n \rangle \otimes \big(g \langle z_n \rangle h \big) \big\} \bar{\eta}_N \langle \bar{z}_0^n \rangle h \\
&\quad + \mu_{N,\bar{b}_0^n} \langle \bar{z}_0^{n-1} \rangle g \langle z_n \rangle^{\otimes 2} \big(\bar{\eta}_N \langle \bar{z}_0^n \rangle h \big)^2.
\end{aligned} \tag{42}$$

Applying Lemma 13 to each term of (42) (note that the second part of Lemma 13 implies the consistency of the updated particle measures, as

$$\bar{\eta}_N \langle \bar{z}_0^n \rangle h = \frac{\eta_N \langle \bar{z}_0^{n-1} \rangle (g \langle z_n \rangle h)}{\eta_N \langle \bar{z}_0^{n-1} \rangle g \langle z_n \rangle} \overset{\mathbb{P}}{\longrightarrow} \frac{\chi \mathbf{Q} \langle \bar{z}_0^{n-1} \rangle (g \langle z_n \rangle h)}{\chi \mathbf{Q} \langle \bar{z}_0^{n-1} \rangle g \langle z_n \rangle} = \bar{\eta} \langle \bar{z}_0^n \rangle h \tag{43}$$

when N tends to infinity, a now classical result) yields, using again (39),

$$\begin{aligned}
\mu_{N,\bar{b}_0^n} &\langle \bar{z}_0^{n-1} \rangle \big\{ g \langle z_n \rangle \big(h - \bar{\eta}_N \langle \bar{z}_0^n \rangle h \big) \big\}^{\otimes 2} \\
&\overset{\mathbb{P}}{\longrightarrow} \mu_{\bar{b}_0^n} \langle \bar{z}_0^{n-1} \rangle \big(g \langle z_n \rangle h \big)^{\otimes 2} \\
&\quad - \mu_{\bar{b}_0^n} \langle \bar{z}_0^{n-1} \rangle \big\{ \big(g \langle z_n \rangle h \big) \otimes g \langle z_n \rangle \big\} \bar{\eta} \langle \bar{z}_0^n \rangle h - \mu_{\bar{b}_0^n} \langle \bar{z}_0^{n-1} \rangle \big\{ g \langle z_n \rangle \otimes \big(g \langle z_n \rangle h \big) \big\} \bar{\eta} \langle \bar{z}_0^n \rangle h \\
&\quad + \mu_{\bar{b}_0^n} \langle \bar{z}_0^{n-1} \rangle g \langle z_n \rangle^{\otimes 2} \big(\bar{\eta} \langle \bar{z}_0^n \rangle h \big)^2 = \mu_{\bar{b}_0^n} \langle \bar{z}_0^{n-1} \rangle \big\{ g \langle z_n \rangle \big(h - \bar{\eta} \langle \bar{z}_0^n \rangle h \big) \big\}^{\otimes 2},
\end{aligned}$$

as N tends to infinity. Now, applying the previous limit to (41) and using (24), (40), and the second part of Lemma 13 yields

$$\bar{\sigma}_{N,\lambda}^2 \langle \bar{z}_0^n \rangle (h) = \frac{\sigma_{N,\lambda}^2 \langle \bar{z}_0^{n-1} \rangle (g \langle z_n \rangle \{ h - \bar{\eta}_N \langle \bar{z}_0^n \rangle h \})}{(\eta_N \langle \bar{z}_0^{n-1} \rangle g \langle z_n \rangle)^2}$$

$$\xrightarrow{\mathbb{P}} \frac{\sigma_\lambda^2 \langle \bar{z}_0^{n-1} \rangle (g \langle z_n \rangle \{h - \bar{\eta} \langle \bar{z}_0^n \rangle h\})}{(\eta \langle \bar{z}_0^{n-1} \rangle g \langle z_n \rangle)^2} = \bar{\sigma}_\lambda^2 \langle \bar{z}_0^n \rangle (h)$$

as N tends to infinity, which completes the proof. □

A.4. Proof of Theorem 9

Proof. Pick arbitrarily $n \in \mathbb{N}$, $\lambda \in \mathbb{N}$, and $h \in \mathbb{F}(\mathcal{X})$. Using the expression of $\bar{\sigma}^2 \langle \bar{\zeta}_0^n \rangle (h)$ derived in the proof of [12], Theorem 11, one may express the bias $\bar{\beta}_\lambda \langle \bar{\zeta}_0^n \rangle (h)$ as

$$\bar{\beta}_\lambda \langle \bar{\zeta}_0^n \rangle (h) = \sum_{m=0}^{n(\lambda)-1} \int \eta \langle \bar{\zeta}_0^{m-1} \rangle (\mathrm{d}x) \left(\frac{\Delta_{\delta_x, \eta \langle \bar{\zeta}_0^{m-1} \rangle} \langle \bar{\zeta}_m^{n-1} \rangle (g \langle \zeta_n \rangle h, g \langle \zeta_n \rangle)}{(\eta \langle \bar{\zeta}_0^{m-1} \rangle \mathbf{Q} \langle \bar{\zeta}_m^{n-1} \rangle \mathbb{1}_\mathsf{X})^2} \right)^2,$$

where $\Delta_{\delta_x, \eta \langle \bar{\zeta}_0^{m-1} \rangle} \langle \bar{\zeta}_m^{n-1} \rangle (g \langle \zeta_n \rangle h, g \langle \zeta_n \rangle)$ is given by (33). Thus, the proof is finalised by following closely the lines of the proof of Theorem 6 and noting that the statement of [12], Proposition 16(iii) still holds true when h and $\mathbb{1}_\mathsf{X}$ are replaced by $g \langle \zeta_n \rangle h$ and $g \langle \zeta_n \rangle$, respectively. □

A.5. Proof of Theorem 10

Proof. If $n \leq \lambda$, the bias vanishes by definition; we thus assume that $n > \lambda$. Write, for $m \in [\![0, n - \lambda]\!]$ and $x \in \mathsf{X}$,

$$\begin{aligned}
&\frac{\mathbf{Q} \langle \bar{z}_m^{n-1} \rangle (h - \eta \langle \bar{z}_0^{n-1} \rangle h)(x)}{\eta \langle \bar{z}_0^{m-1} \rangle \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle \mathbb{1}_\mathsf{X}} \\
&\qquad = \frac{\delta_x \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle \mathbb{1}_\mathsf{X}}{\eta \langle \bar{z}_0^{m-1} \rangle \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle \mathbb{1}_\mathsf{X}} \left(\frac{\delta_x \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle h}{\delta_x \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle \mathbb{1}_\mathsf{X}} - \frac{\eta \langle \bar{z}_0^{m-1} \rangle \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle h}{\eta \langle \bar{z}_0^{m-1} \rangle \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle \mathbb{1}_\mathsf{X}} \right),
\end{aligned} \tag{44}$$

where [5], Proposition 4.3.4 bounds uniformly the second factor of (44) according to

$$\left| \frac{\delta_x \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle h}{\delta_x \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle \mathbb{1}_\mathsf{X}} - \frac{\eta \langle \bar{z}_0^{m-1} \rangle \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle h}{\eta \langle \bar{z}_0^{m-1} \rangle \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle \mathbb{1}_\mathsf{X}} \right| \leq \varrho^{n-m} \|h\|_\infty. \tag{45}$$

To bound the first factor of (44), note that

$$\eta \langle \bar{z}_0^{m-1} \rangle \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle \mathbb{1}_\mathsf{X} = \bar{\eta} \langle \bar{z}_0^{m-1} \rangle \mathbf{M} (g \langle z_m \rangle \mathbf{M} \mathbf{Q} \langle \bar{z}_{m+1}^{n-1} \rangle \mathbb{1}_\mathsf{X}) \geq \underline{\delta} \varepsilon \mu \mathbf{Q} \langle \bar{z}_{m+1}^{n-1} \rangle \mathbb{1}_\mathsf{X}$$

and

$$\delta_x \mathbf{Q} \langle \bar{z}_m^{n-1} \rangle \mathbb{1}_\mathsf{X} = g \langle z_m \rangle (x) \mathbf{M} \mathbf{Q} \langle \bar{z}_{m+1}^{n-1} \rangle \mathbb{1}_\mathsf{X}(x) \leq \bar{\delta} \bar{\varepsilon} \mu \mathbf{Q} \langle \bar{z}_{m+1}^{n-1} \rangle \mathbb{1}_\mathsf{X},$$

where $(\underline{\varepsilon}, \bar{\varepsilon})$ and $(\underline{\delta}, \bar{\delta})$ are given in **(S)**(ii), implying that

$$\frac{\delta_x \mathbf{Q}\langle \bar{z}_m^{n-1}\rangle \mathbb{1}_{\mathsf{X}}}{\eta\langle \bar{z}_0^{m-1}\rangle \mathbf{Q}\langle \bar{z}_m^{n-1}\rangle \mathbb{1}_{\mathsf{X}}} \le \frac{\bar{\delta}\bar{\varepsilon}}{\underline{\delta}\underline{\varepsilon}}. \tag{46}$$

Now, using (45) and (46), proceed like

$$\beta_\lambda\langle \bar{z}_0^{n-1}\rangle(h) = \sum_{m=0}^{n-\lambda-1} \frac{\eta\langle \bar{z}_0^{m-1}\rangle\{\mathbf{Q}\langle \bar{z}_m^{n-1}\rangle(h - \eta\langle \bar{z}_0^{n-1}\rangle h)\}^2}{(\eta\langle \bar{z}_0^{m-1}\rangle \mathbf{Q}\langle \bar{z}_m^{n-1}\rangle \mathbb{1}_{\mathsf{X}})^2}$$

$$\le \|h\|_\infty^2 \left(\frac{\bar{\delta}\bar{\varepsilon}}{\underline{\delta}\underline{\varepsilon}}\right)^2 \sum_{m=0}^{n-\lambda-1} \varrho^{2(n-m)} \le c\|h\|_\infty^2 \varrho^{2(\lambda+1)},$$

with $c := (\bar{\delta}\bar{\varepsilon}/\underline{\delta}\underline{\varepsilon})^2/(1 - \varrho^2)$, and the proof is complete. $\qquad\square$

Acknowledgments

References

[1] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. *Springer Series in Statistics*. New York: Springer. MR2159833

[2] Chan, H.P. and Lai, T.L. (2013). A general theory of particle filters in hidden Markov models and some applications. *Ann. Statist.* **41** 2877–2904. MR3161451

[3] Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. MR2153989

[4] Crisan, D. and Heine, K. (2008). Stability of the discrete time filter in terms of the tails of noise distributions. *J. Lond. Math. Soc.* (2) **78** 441–458. MR2439634

[5] Del Moral, P. (2004). *Feynman–Kac Formulae*: *Genealogical and Interacting Particle Systems with Applications*. *Probability and Its Applications* (*New York*). New York: Springer. MR2044973

[6] Del Moral, P. (2013). *Mean Field Simulation for Monte Carlo Integration*. *Monographs on Statistics and Applied Probability* **126**. Boca Raton, FL: CRC Press. MR3060209

[7] Del Moral, P. and Guionnet, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* **9** 275–297. MR1687359

[8] Del Moral, P. and Guionnet, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 155–194. MR1819122

[9] Douc, R., Fort, G., Moulines, E. and Priouret, P. (2009). Forgetting the initial distribution for hidden Markov models. *Stochastic Process. Appl.* **119** 1235–1256. MR2508572

[10] Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376. MR2458190

[11] Douc, R. and Moulines, E. (2012). Asymptotic properties of the maximum likelihood estimation in misspecified hidden Markov models. *Ann. Statist.* **40** 2697–2732. MR3097617

[12] Douc, R., Moulines, E. and Olsson, J. (2014). Long-term stability of sequential Monte Carlo methods under verifiable conditions. *Ann. Appl. Probab.* **24** 1767–1802. MR3226163

[13] Doucet, A., De Freitas, N. and Gordon, N., eds. (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer.

[14] Gordon, N., Salmond, D. and Smith, A.F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc., F, Radar Signal Process.* **140** 107–113.

[15] Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance* **42** 281–300.

[16] Jacob, P.E., Murray, L.M. and Rubenthaler, S. (2015). Path storage in the particle filter. *Stat. Comput.* **25** 487–496. MR3306720

[17] Kitagawa, G. and Sato, S. (2001). Monte Carlo smoothing and self-organising state-space model. In *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. De Freitas and N. Gordon, eds.). *Stat. Eng. Inf. Sci.* 177–195. New York: Springer. MR1847792

[18] Künsch, H.R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. MR2211077

[19] Lee, A. and Whiteley, N. (2016). Variance estimation in particle filters. Preprint. Available at arXiv:1509.00394.

[20] Lindsten, F., Schön, T.B. and Olsson, J. (2011). An explicit variance reduction expression for the Rao–Blackwellised particle filter. In *Proceedings of the* 18*th IFAC World Congress* 11979–11984.

[21] Olsson, J., Cappé, O., Douc, R. and Moulines, E. (2008). Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models. *Bernoulli* **14** 155–179. MR2401658

[22] Olsson, J. and Ströjby, J. (2011). Particle-based likelihood inference in partially observed diffusion processes using generalised Poisson estimators. *Electron. J. Stat.* **5** 1090–1122. MR2836770

[23] Ristic, B., Arulampalam, M. and Gordon, A. (2004). *Beyond Kalman Filters*: *Particle Filters for Target Tracking*. Norwood: Artech House.