Bernoulli **22**(1), 2016, 494–529
DOI: 10.3150/14-BEJ666

On the role of interaction in sequential Monte Carlo algorithms

NICK WHITELEY[1,*], ANTHONY LEE[2,**] and KARI HEINE[1,†]

[1]*School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, United Kingdom.*
E-mail: [*]*nick.whiteley@bristol.ac.uk*; [†]*kari.heine@bristol.ac.uk*
[2]*Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom.*
E-mail: [**]*anthony.lee@warwick.ac.uk*

We introduce a general form of sequential Monte Carlo algorithm defined in terms of a parameterized resampling mechanism. We find that a suitably generalized notion of the Effective Sample Size (ESS), widely used to monitor algorithm degeneracy, appears naturally in a study of its convergence properties. We are then able to phrase sufficient conditions for time-uniform convergence in terms of algorithmic control of the ESS, in turn achievable by adaptively modulating the interaction between particles. This leads us to suggest novel algorithms which are, in senses to be made precise, provably stable and yet designed to avoid the degree of interaction which hinders parallelization of standard algorithms. As a byproduct, we prove time-uniform convergence of the popular adaptive resampling particle filter.

Keywords: convergence; hidden Markov model; particle filters

1. Introduction

At the frontier of computational statistics there is growing interest in parallel implementation of Monte Carlo algorithms using multi-processor and distributed architectures. However, the resampling step of sequential Monte Carlo (SMC) methods [13] (see [16] for a recent overview) which involves a degree of interaction between simulated "particles", hinders their parallelization. So, whilst multi-processor implementation offers some speed up for SMC, the potential benefits of distributed computing are not fully realized [17].

Performing resampling only occasionally, a technique originally suggested for the somewhat different reason of variance reduction [19], alleviates this problem to some extent, but the collective nature of the resampling operation remains the computational bottleneck. On the other hand, crude attempts to entirely do away with the resampling step may result in unstable or even non-convergent algorithms. With these issues in mind, we seek a better understanding of the relationship between the interaction structure of SMC algorithms and theoretical properties of the approximations they deliver. Our overall aim is to address the following question:

To what extent can the degree of interaction between particles be reduced, whilst ensuring provable stability of the algorithm?

Our strategy is to introduce and study an unusually general type of SMC algorithm featuring a parameterized resampling mechanism. This provides a flexible framework in which we are

ultimately able to attach meaning to *degree of interaction* in terms of graph-theoretic quantities. To address the matter of *provable stability*, we seek conditions under which the algorithm yields time-uniformly convergent approximations of prediction filters, and approximations of marginal likelihoods whose relative variance can be controlled at a linear-in-time cost.

The general algorithm we study is defined in terms of a family of Markov transition matrices, α, and we refer to the algorithm itself as αSMC. We shall see that through particular choices of α one obtains, as instances of αSMC, well-known algorithms including sequential importance sampling (SIS), the bootstrap particle filter (BPF) and the adaptive resampling particle filter (ARPF) in which resampling is triggered by monitoring some functional criterion, such as the Effective Sample Size (ESS) [19].

Although the ESS does not necessarily appear in the definition of the general αSMC algorithm, we find that it does appear quite naturally from the inverse quadratic variation of certain martingale sequences in its analysis. This allows us to make precise a sense in which algorithmic control of the ESS can guarantee stability of the algorithm. Our results apply immediately to the ARPF, but our study has wider-reaching methodological consequences: in our framework it becomes clear that the standard adaptive resampling strategy is just one of many possible ways of algorithmically controlling the ESS, and we can immediately suggest new, alternative algorithms which are provably stable, but designed to avoid the type of complete interaction which is inherent to the ARPF and which hinders its parallelization. The structure of this paper and our main contributions are as follows.

Section 2 introduces the general algorithm, αSMC. We explain how it accommodates several standard algorithms as particular cases and comment on some other existing SMC methods.

Section 3 presents Theorem 1, a general convergence result for αSMC. We give conditions which ensure unbiased approximation of marginal likelihoods and we elucidate connections between certain invariance properties of the matrices α and the negligibility of increments in a martingale error decomposition, thus formulating simple sufficient conditions for weak and strong laws of large numbers. We also discuss some related existing results.

Section 4 presents our second main result, Theorem 2. We show, subject to regularity conditions on the hidden Markov model (HMM) under consideration, that enforcement of a strictly positive lower bound on a certain coefficient associated with ESS of αSMC is sufficient to guarantee non-asymptotic, time-uniform bounds on: (1) the exponentially normalized relative second moment of error in approximation of marginal likelihoods, and (2) the L_p norm of error in approximation of prediction filters. The former implies a linear-in-time variance bound and the latter implies time-uniform convergence. These results apply immediately to the ARPF.

Section 5 houses discussion and application of our results. We point out the pitfalls of some naive approaches to parallelization of SMC and discuss what can go wrong if the conditions of Theorem 1 are not met. Three new algorithms, which adapt the degree of interaction in order to control the ESS and which are therefore provably stable, are then introduced. We discuss computational complexity and through numerical experiments examine the degree of interaction involved in these algorithms and the quality of the approximations they deliver compared to the ARPF.

2. αSMC

A hidden Markov model (HMM) with measurable state space $(\mathsf{X}, \mathcal{X})$ and observation space $(\mathsf{Y}, \mathcal{Y})$ is a process $\{(X_n, Y_n); n \geq 0\}$ where $\{X_n; n \geq 0\}$ is a Markov chain on X, and each observation Y_n, valued in Y, is conditionally independent of the rest of the process given X_n. Let μ_0 and f be, respectively, a probability distribution and a Markov kernel on $(\mathsf{X}, \mathcal{X})$, and let g be a Markov kernel acting from $(\mathsf{X}, \mathcal{X})$ to $(\mathsf{Y}, \mathcal{Y})$, with $g(x, \cdot)$ admitting a density, denoted similarly by $g(x, y)$, with respect to some dominating σ-finite measure. The HMM specified by μ_0, f and g, is

$$\begin{aligned} X_0 \sim \mu_0(\cdot), \qquad X_n|\{X_{n-1} = x_{n-1}\} \sim f(x_{n-1}, \cdot), \qquad n \geq 1, \\ Y_n|\{X_n = x_n\} \sim g(x_n, \cdot), \qquad n \geq 0. \end{aligned} \tag{1}$$

We shall assume throughout that we are presented with a fixed observation sequence $\{y_n; n \geq 0\}$ and write

$$g_n(x) := g(x, y_n), \qquad n \geq 0.$$

The following assumption imposes some mild regularity which ensures that various objects appearing below are well defined. It shall be assumed to hold throughout without further comment.

Assumption. (A1) *For each* $n \geq 0$, $\sup_x g_n(x) < +\infty$ *and* $g_n(x) > 0$ *for all* $x \in \mathsf{X}$.

We take as a recursive definition of the *prediction filters*, the sequence of distributions $\{\pi_n; n \geq 0\}$ given by

$$\begin{aligned} \pi_0 &:= \mu_0, \\ \pi_n(A) &:= \frac{\int_{\mathsf{X}} \pi_{n-1}(\mathrm{d}x) g_{n-1}(x) f(x, A)}{\int_{\mathsf{X}} \pi_{n-1}(\mathrm{d}x) g_{n-1}(x)}, \qquad A \in \mathcal{X}, n \geq 1, \end{aligned} \tag{2}$$

and let $\{Z_n; n \geq 0\}$ be defined by

$$Z_0 := 1, \qquad Z_n := Z_{n-1} \int_{\mathsf{X}} \pi_{n-1}(\mathrm{d}x) g_{n-1}(x), \qquad n \geq 1. \tag{3}$$

Due to the conditional independence structure of the HMM, π_n is the conditional distribution of X_n given $Y_{0:n-1} = y_{0:n-1}$; and Z_n is the marginal likelihood of the first n observations, evaluated at the point $y_{0:n-1}$. Our main computational objectives are to approximate $\{\pi_n; n \geq 0\}$ and $\{Z_n; n \geq 0\}$.

2.1. The general algorithm

With population size $N \geq 1$, we write $[N] := \{1, \ldots, N\}$. *To simplify presentation, whenever a summation sign appears without the summation set made explicit, the summation set is taken to be* $[N]$*, for example, we write* $\sum_i$ *to mean* $\sum_{i=1}^N$.

The αSMC algorithm involves simulating a sequence $\{\zeta_n; n \geq 0\}$ with each $\zeta_n = \{\zeta_n^1, \ldots, \zeta_n^N\}$ valued in X^N. Denoting $\mathbb{X} := (\mathsf{X}^N)^{\mathbb{N}}$, $\mathcal{F}^{\mathbb{X}} := (\mathcal{X}^{\otimes N})^{\otimes \mathbb{N}}$, we shall view $\{\zeta_n; n \geq 0\}$ as the canonical coordinate process on the measurable space $(\mathbb{X}, \mathcal{F}^{\mathbb{X}})$, and write $\mathcal{F}_n$ for the σ-algebra generated by $\{\zeta_0, \ldots, \zeta_n\}$. By convention, we let $\mathcal{F}_{-1} := \{\mathbb{X}, \varnothing\}$ be the trivial σ-algebra. The sampling steps of the αSMC algorithm, described below, amount to specifying a probability measure, say $\mathbb{P}$, on $(\mathbb{X}, \mathcal{F}^{\mathbb{X}})$. Expectation w.r.t. $\mathbb{P}$ shall be denoted by $\mathbb{E}$.

Let $\mathbb{A}_N$ be a non-empty set of Markov transition matrices, each of size $N \times N$. For $n \geq 0$ let $\alpha_n : \mathbb{X} \to \mathbb{A}_N$ be a matrix-valued map, and write α_n^{ij} for the ith row, jth column entry so that for each i we have $\sum_j \alpha_n^{ij} = 1$ (with dependence on the $\mathbb{X}$-valued argument suppressed). The following assumption places a restriction on the relationship between α and the particle system $\{\zeta_n; n \geq 0\}$.

Assumption. (A2) *For each $n \geq 0$, the entries of α_n are all measurable with respect to $\mathcal{F}_n$*

Intuitively, the members of $\mathbb{A}_N$ will specify different possible interaction structures for the particle algorithm and under (A2), each α_n is a random matrix chosen from $\mathbb{A}_N$ according to some deterministic function of $\{\zeta_0, \ldots, \zeta_n\}$. Examples are given below. We shall write $\mathbf{1}_{1/N}$ for the $N \times N$ matrix which has $1/N$ as every entry and write Id for the identity matrix of size apparent from the context in which this notation appears. We shall occasionally use Id also to denote identity operators in certain function space settings. Let $\mathcal{M}$, $\mathcal{P}$ and $\mathcal{L}$ be, respectively, the collections of measures, probability measures and real-valued, bounded, $\mathcal{X}$-measurable functions on X. We write

$$\|\varphi\| := \sup_x \bigl|\varphi(x)\bigr|, \qquad \operatorname{osc}(\varphi) := \sup_{x,y} \bigl|\varphi(x) - \varphi(y)\bigr|$$

and

$$\mu(\varphi) := \int_{\mathsf{X}} \varphi(x)\mu(\mathrm{d}x) \qquad \text{for any } \varphi \in \mathcal{L}, \mu \in \mathcal{M}. \tag{4}$$

Remark. Note that $\mathbb{X}$, $\mathcal{F}^{\mathbb{X}}$, $\mathcal{F}_n$, $\mathbb{P}$, α and various other objects depend on N, but this dependence is suppressed from the notation. Unless specified otherwise, any conditions which we impose on such objects should be understood as holding for all $N \geq 1$.

Let $\{W_n^i; i \in [N], n \geq 0\}$ be defined by the following recursion:

$$W_0^i := 1, \qquad W_n^i := \sum_j \alpha_{n-1}^{ij} W_{n-1}^j g_{n-1}\bigl(\zeta_{n-1}^j\bigr), \qquad i \in [N], n \geq 1. \tag{5}$$

The following algorithm implicitly specifies the law $\mathbb{P}$ of the αSMC particle system. For each $n \geq 1$, the "Sample" step should be understood as meaning that the variables $\zeta_n = \{\zeta_n^i\}_{i \in [N]}$ are conditionally independent given $\{\zeta_0, \ldots, \zeta_{n-1}\}$. The line of Algorithm 1 marked ($\star$) is intentionally generic, it amounts to a practical, if imprecise restatement of (A2). In the sequel, we shall examine instances of αSMC which arise when we consider specific $\mathbb{A}_N$ and impose more structure at line ($\star$).

Algorithm 1 αSMC

For $n = 0$,
  For $i = 1, \ldots, N$,
    Set $W_0^i = 1$
    Sample $\zeta_0^i \sim \mu_0$
For $n \geq 1$,
($\star$) Select α_{n-1} from $\mathbb{A}_N$ according to some functional of $\{\zeta_0, \ldots, \zeta_{n-1}\}$.
  For $i = 1, \ldots, N$,
    Set $W_n^i = \sum_j \alpha_{n-1}^{ij} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j)$.
    Sample $\zeta_n^i | \mathcal{F}_{n-1} \sim \frac{\sum_j \alpha_{n-1}^{ij} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j) f(\zeta_{n-1}^j, \cdot)}{W_n^i}$.

We shall study the objects

$$\pi_n^N := \frac{\sum_i W_n^i \delta_{\zeta_n^i}}{\sum_i W_n^i}, \qquad Z_n^N := \frac{1}{N} \sum_i W_n^i, \qquad n \geq 0, \tag{6}$$

which as the notation suggests, are to be regarded as approximations of π_n and Z_n, respectively. We shall also be centrally concerned with the following coefficient, which is closely related to the ESS,

$$\mathcal{E}_n^N := \frac{(N^{-1} \sum_i W_n^i)^2}{N^{-1} \sum_i (W_n^i)^2} = \frac{(N^{-1} \sum_i \sum_j \alpha_{n-1}^{ij} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j))^2}{N^{-1} \sum_i (\sum_j \alpha_{n-1}^{ij} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j))^2}, \qquad n \geq 1, \tag{7}$$

and by convention $\mathcal{E}_0^N := 1$. The second equality in (7) is immediate from the definition of W_n^i, see (5). Note that $\mathcal{E}_n^N$ is always valued in [0, 1], and if we write

$$N_n^{\text{eff}} := N \mathcal{E}_n^N, \tag{8}$$

we obtain the ESS of Liu and Chen [19], although of course in a generalized form, since $\mathcal{E}_n^N$ is defined in terms of the generic ingredients of αSMC. A few comments on generality are in order. First, for ease of presentation, we have chosen to work with a particularly simple version of αSMC, in which new samples are proposed using the HMM Markov kernel f. The algorithm is easily generalized to accommodate other proposal kernels. Second, while we focus on the application of SMC methods to HMM's, our results and methodological ideas are immediately transferable to other contexts, for example, via the framework of [10].

2.2. Instances of αSMC

We now show how αSMC admits SIS, the BPF and the ARPF, as special cases, through particular choices of $\mathbb{A}_N$. Our presentation is intended to illustrate the structural generality of αSMC, thus

setting the scene for the developments which follow. The following lemma facilitates exposition by "unwinding" the quantities $\{W_n^i\}_{i\in[N]}$ defined recursively in (5). It is used throughout the remainder of the paper.

Lemma 1. *For $n \geq 1$, $0 \leq p < n$ and $i_n \in [N]$,*

$$W_n^{i_n} = \sum_{(i_p,\ldots,i_{n-1})\in[N]^{n-p}} W_p^{i_p} \prod_{q=p}^{n-1} g_q\left(\zeta_q^{i_q}\right)\alpha_q^{i_{q+1}i_q}, \tag{9}$$

and in particular

$$W_n^{i_n} = \sum_{(i_0,\ldots,i_{n-1})\in[N]^n} \prod_{p=0}^{n-1} g_p\left(\zeta_p^{i_p}\right)\alpha_p^{i_{p+1}i_p}. \tag{10}$$

The proof of (9)–(10) is a simple induction and is therefore omitted. From (10) and definitions above, we immediately observe the following corollary.

Corollary 1. *If* (A2) *holds, then W_n^i must be measurable w.r.t. $\mathcal{F}_{n-1}$ for every $n \geq 0$ and $i \in [N]$.*

Sequential importance sampling: $\mathbb{A}_N = \{\mathrm{Id}\}$

Since in this case $\mathbb{A}_N$ consists of only a single element, α is actually a deterministic sequence, (A2) is trivially satisfied and at line ($\star$) of Algorithm 1 we have $\alpha_n = \mathrm{Id}$ fixed for all $n \geq 0$. In this situation, Lemma 1 gives $W_n^i = \prod_{p=0}^{n-1} g_p(\zeta_p^i)$ for $n \geq 1$, so in turn

$$\pi_n^N = \frac{\sum_i \delta_{\zeta_n^i} \prod_{p=0}^{n-1} g_p(\zeta_p^i)}{\sum_i \prod_{p=0}^{n-1} g_p(\zeta_p^i)}, \qquad Z_n^N = \frac{1}{N}\sum_i \prod_{p=0}^{n-1} g_p\left(\zeta_p^i\right), \qquad n \geq 1,$$

and αSMC reduces to Algorithm 2.

Algorithm 2 Sequential importance sampling

For $n = 0$,
  For $i = 1, \ldots, N$,
    Set $W_0^i = 1$
    Sample $\zeta_0^i \sim \mu_0$
For $n \geq 1$,
  For $i = 1, \ldots, N$,
    Set $W_n^i = W_{n-1}^i g_{n-1}(\zeta_{n-1}^i)$
    Sample $\zeta_n^i | \mathcal{F}_{n-1} \sim f(\zeta_{n-1}^i, \cdot)$

Bootstrap particle filter: $\mathbb{A}_N = \{\mathbf{1}_{1/N}\}$

In this case, α is again a deterministic sequence and (A2) is trivially satisfied. At line ($\star$) we have $\alpha_n = \mathbf{1}_{1/N}$ fixed for all $n \geq 0$. Lemma 1 gives, for all $i_n \in [N]$,

$$W_n^{i_n} = \sum_{(i_0,\ldots,i_{n-1})\in[N]^n} \prod_{p=0}^{n-1} \frac{g_p(\zeta_p^{i_p})}{N} = \prod_{p=0}^{n-1}\left(\frac{1}{N}\sum_{i_p} g_p\left(\zeta_p^{i_p}\right)\right), \qquad n \geq 1. \tag{11}$$

Note that then $W_n^i = W_n^j$ for all i, j, so $NW_n^i = \sum_j W_n^j$ and we obtain, according to (6),

$$\pi_n^N = \frac{1}{N}\sum_i \delta_{\zeta_n^i}, \qquad Z_n^N = \prod_{p=0}^{n-1}\left(\frac{1}{N}\sum_{i_p} g_p\left(\zeta_p^{i_p}\right)\right), \qquad n \geq 1, \tag{12}$$

and αSMC algorithm reduces to Algorithm 3. Since $W_n^i = W_n^j$ for all i, j, we write by convention the weight update steps only for W_n^1.

Adaptive resampling particle filter: $\mathbb{A}_N = \{\mathrm{Id}, \mathbf{1}_{1/N}\}$

In this case, each α_n is allowed to take only the value Id or $\mathbf{1}_{1/N}$, with the latter corresponding to resampling, and vice-versa. The choice between Id and $\mathbf{1}_{1/N}$ is made by comparing some functional of the particle system to a threshold value. We consider the case of the popular ESS-based resampling rule [19], partly for simplicity, but also because monitoring of the ESS is especially pertinent to the discussions which follow. This ARPF arises as an instance of αSMC if we take as line ($\star$) of Algorithm 1 the rule:

$$\alpha_{n-1} := \begin{cases} \mathbf{1}_{1/N}, & \text{if } \dfrac{(N^{-1}\sum_i W_{n-1}^i g_{n-1}(\zeta_{n-1}^i))^2}{N^{-1}\sum_i (W_{n-1}^i g_{n-1}(\zeta_{n-1}^i))^2} < \tau, \\ \mathrm{Id}, & \text{otherwise,} \end{cases} \tag{13}$$

where $\tau \in (0, 1]$ is a threshold value. Lemma 4 in the Appendix shows by an inductive argument that the adaptation rule (13) satisfies (A2). The ARPF is traditionally expressed in terms of the random times at which resampling occurs. For completeness, the Appendix contains derivations

Algorithm 3 Bootstrap particle filter

For $n = 0$,
  Set $W_0^1 = 1$
  For $i = 1, \ldots, N$,
    Sample $\zeta_0^i \sim \mu_0$
For $n \geq 1$,
  Set $W_n^1 = W_{n-1}^1 \cdot \frac{1}{N}\sum_i g_{n-1}(\zeta_{n-1}^i)$
  For $i = 1, \ldots, N$,
    Sample $\zeta_n^i | \mathcal{F}_{n-1} \sim \dfrac{\sum_j g_{n-1}(\zeta_{n-1}^j) f(\zeta_{n-1}^j, \cdot)}{\sum_j g_{n-1}(\zeta_{n-1}^j)}$

of expressions for π_n^N and Z_n^N in terms of such times and similar manipulations can be used to write out the form of αSMC in this case.

Looking back to the expression for $\mathcal{E}_n^N$ in (7), we find:

$$\alpha_{n-1} = \mathbf{1}_{1/N} \quad \Rightarrow \quad \mathcal{E}_n^N = 1, \tag{14}$$

$$\alpha_{n-1} = \mathrm{Id} \quad \Rightarrow \quad \mathcal{E}_n^N = \frac{(N^{-1}\sum_i W_{n-1}^i g_{n-1}(\zeta_{n-1}^i))^2}{N^{-1}\sum_i (W_{n-1}^i g_{n-1}(\zeta_{n-1}^i))^2}. \tag{15}$$

We then adopt the point of view that according to (13)–(15), the ARPF *enforces* the condition: $\inf_{n\geq 0} \mathcal{E}_n^N \geq \tau > 0$, or equivalently

$$\inf_{n\geq 0} N_n^{\mathrm{eff}} \geq N\tau > 0,$$

by construction. This seemingly trivial observation turns out to be crucial when we address time-uniform convergence of the ARPF in Section 4, and the condition $\inf_{n\geq 0} \mathcal{E}_n^N > 0$ will appear repeatedly in discussions which lead to the formulation of new, provably stable algorithms in Section 5.

To give some flavour of the kind of algorithms we have in mind, let $(B_L^\ell)_{\ell=1,\ldots,L}$ be a partition of the set $[N]$ into L clusters, and suppose the matrix α_{n-1} is defined by $\alpha_{n-1}^{ij} = 1/|B_L^\ell|$ if both $i, j \in B_L^\ell$ and 0 otherwise. Then for any $\ell = 1, \ldots, L$ and any $i \in B_L^\ell$ the new weight W_n^i and the distribution from which the new particle ζ_n^i is sampled, say m_n^i (both of which depend on ℓ only), are given by

$$W_n^i = \frac{1}{|B_L^\ell|} \sum_{j\in B_L^\ell} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j), \qquad m_n^i = \frac{\sum_{j\in B_L^\ell} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j) f(\zeta_{n-1}^j, \cdot)}{\sum_{j\in B_L^\ell} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j)}. \tag{16}$$

Furthermore, in this situation we have

$$\mathcal{E}_n^N = \frac{(N^{-1}\sum_i W_{n-1}^i g_{n-1}(\zeta_{n-1}^i))^2}{\sum_{\ell=1}^L |B_L^\ell|/N(|B_L^\ell|^{-1}\sum_{j\in B_L^\ell} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j))^2}.$$

It is then clear that a parallel implementation could be possible, say on L processors, one devoted to each cluster and it remains to design an efficient partition of $(B_L^\ell)_{\ell=1,\ldots,L}$ of the set $[N]$.

Comments on other algorithms

In the engineering literature, a variety of algorithmic procedures involving distributed computing have been suggested [3], including partitioning ideas like (16). "Local" particle approximations of Rao–Blackwellized filters have been devised in [6] and [14]. Vergé *et al.* [21] have recently suggested an "island" particle algorithm, designed for parallel implementation, in which there are two levels of resampling and the total population size $N = N_1 N_2$ is defined in terms of the number of particles per island, N_1, and the number of islands, N_2. Interaction at both levels occurs by resampling, at the island level this means entire blocks of particles are replicated

and/or discarded. They investigate the trade-off between N_1 and N_2 and provide asymptotic results which validate their algorithms. In the present work, we provide some asymptotic results in Section 3 but it is really the non-asymptotic results in Section 4 which lead us to suggest specific novel instances of αSMC in Section 5. Moreover, in general αSMC is distinct from all these algorithms and, other than in some uninteresting special cases, none of them coincide with the adaptive procedures we suggest in Section 5.3.

3. Convergence

In this section, our main objective is to investigate, for general αSMC (Algorithm 1), conditions for convergence

$$Z_n^N - Z_n \to 0 \quad \text{and} \quad \pi_n^N(\varphi) - \pi_n(\varphi) \to 0, \tag{17}$$

at least in probability, as $N \to \infty$.

In the case of SIS, that is, $\mathbb{A}_N = \{\mathrm{Id}\}$, it is easy to establish (17), since the processes $\{\zeta_n^i; n \geq 0\}_{i\in[N]}$ are independent Markov chains, of identical law. On the other hand, for the bootstrap filter, that is, $\mathbb{A}_N = \{\mathbf{1}_{1/N}\}$, the convergence $\pi_n^N(\varphi) - \pi_n(\varphi) \to 0$, can be proved under very mild conditions, by decomposing $\pi_n^N(\varphi) - \pi_n(\varphi)$ in terms of "local" sampling errors, see amongst others [8,12] for this type of approach. For instance, for $A \in \mathcal{X}$ we may write

$$\pi_1^N(A) - \pi_1(A) = \frac{1}{N}\sum_i \delta_{\zeta_1^i}(A) - \frac{\sum_i g_0(\zeta_0^i) f(\zeta_0^i, A)}{\sum_i g_0(\zeta_0^i)} \tag{18}$$

$$+ \frac{\sum_i g_0(\zeta_0^i) f(\zeta_0^i, A)}{\sum_i g_0(\zeta_0^i)} - \pi_1(A). \tag{19}$$

Heuristically, the term on the r.h.s. of (18) converges to zero because given $\mathcal{F}_0$, the samples $\{\zeta_1^i\}_{i\in[N]}$ are conditionally i.i.d. according $\frac{\sum_i g_0(\zeta_0^i) f(\zeta_0^i, \cdot)}{\sum_i g_0(\zeta_0^i)}$, and the term in (19) converges to zero because the samples $\{\zeta_0^i\}_{i\in[N]}$ are i.i.d. according to μ_0. A similar argument ensures that $\pi_n^N(\varphi) - \pi_n(\varphi) \to 0$, for any $n \geq 0$ and therefore by the continuous mapping theorem $Z_n^N - Z_n \to 0$, since

$$Z_n = \prod_{p=0}^{n-1} \pi_p(g_p) \quad \text{and} \quad Z_n^N = \prod_{p=0}^{n-1} \pi_p^N(g_p).$$

In the case of αSMC, $\{\zeta_n^i\}_{i\in[N]}$ are conditionally independent given $\mathcal{F}_{n-1}$, but we do not necessarily have either the unconditional independence structure of SIS, or the conditionally i.i.d. structure of the BPF to work with.

Douc and Moulines [12] have established a CLT for the ARPF using an inductive approach w.r.t. deterministic time periods. Arnaud and Le Gland [2] have obtained a CLT for the ARPF based on an alternative multiplicative functional representation of the algorithm. Convergence of the ARPF was studied in [11] by coupling the adaptive algorithm to a reference particle system, for which resampling occurs at deterministic times. One of the benefits of their approach is that

existing asymptotic results for non-adaptive algorithms, such as central limit theorems (CLT), can then be transferred to the adaptive algorithm with little further work. Their analysis involves a technical assumption [11], Section 5.2, to deal with the situation where the threshold parameters coincide with the adaptive criteria. Our analysis of αSMC does not rest on any such technical assumption, and in some ways is more direct, but we do not obtain concentration estimates or a CLT. Some more detailed remarks on this matter are given after the statement of Theorem 1.

Crisan and Obanubi [7] studied convergence and obtained a CLT for an adaptive resampling particle filter in continuous time under conditions which they verify for the case of ESS-triggered resampling, without needing the type of technical assumption of [11]. Their study focuses, in part, on the random times at which resampling occurs and dealing with the subtleties of the convergence in continuous time. Our asymptotic $N \to \infty$ analysis is in some ways less refined, but in comparison to this and the other existing works, we analyze a more general algorithm, and it is this generality which allows us to suggest new adaptive algorithms in Section 5, informed by the time-uniform non-asymptotic error bounds in our Theorem 2.

To proceed, we need some further notation involving α. Let us define the matrices: $\alpha_{n,n} := \mathrm{Id}$ for $n \geq 0$, and recursively

$$\alpha_{p,n}^{ij} := \sum_k \alpha_{p+1,n}^{ik} \alpha_p^{kj}, \qquad (i,j) \in [N]^2, 0 \leq p < n, \tag{20}$$

and the vectors:

$$\beta_{n,n}^i := N^{-1}, \qquad n \geq 0, i \in [N], \tag{21}$$

and recursively

$$\beta_{p,n}^i := \sum_j \beta_{p+1,n}^j \alpha_p^{ji}, \qquad i \in [N], 0 \leq p < n. \tag{22}$$

Note that since each α_n is a random Markov transition matrix, so is each $\alpha_{p,n}$, and each $\{\beta_{p,n}^i\}_{i\in[N]}$ defines a random probability distribution on $[N]$. Moreover, from these definitions we immediately have the identity

$$\beta_{p,n}^i = N^{-1} \sum_j \alpha_{p,n}^{ji}, \qquad i \in [N], 0 \leq p \leq n. \tag{23}$$

Assumption. (B) – *for all* $0 \leq p \leq n$ *and* $i \in [N]$, $\beta_{p,n}^i$ *is measurable w.r.t. the trivial* σ*-algebra* $\mathcal{F}_{-1}$.

(B$^+$) – *assumption* (B) *holds and, for all* $0 \leq p \leq n$, $\lim_{N\to\infty} \max_{i\in[N]} \beta_{p,n}^i = 0$.

(B^{++}) – *every member of* $\mathbb{A}_N$ *admits the uniform distribution on* $[N]$ *as an invariant distribution*

We note the following:

- Intuitively, (B) ensures that even though α is a sequence of random Markov transition matrices, the elements of the probability vector $\{\beta_{p,n}^i\}_{i\in[N]}$ are all constants. (B) holds, trivially,

when every element of every α_n is measurable w.r.t. $\mathcal{F}_{-1}$, that is, the sequence α is completely pre-determined. This is true, for example, when the set $\mathbb{A}_N$ consists of only a single element, as is the case for SIS and the BPF.

- The $\lim_{N\to\infty} \max_{i\in[N]} \beta^i_{p,n} = 0$ part of (B^+) is an asymptotic negligibility condition. In Section 5.2, we describe what can go wrong when this assumption does not hold.
- (B^{++}) does not require the members of $\mathbb{A}_N$ to be irreducible, for example, it is satisfied with $\mathbb{A}_N = \{\mathrm{Id}\}$.
- (B^{++}) $\Rightarrow$ (B^+). To see this, observe that when (B^{++}) holds, every random matrix $\alpha_{p,n}$, defined in (20), also admits the uniform distribution on $[N]$ as invariant, then using (23) we have $\beta^i_{p,n} = N^{-1}\sum_j \alpha^{ji}_{p,n} = N^{-1}$ for all $i \in [N]$. The reverse implication is clearly not true in general.
- (B^{++}) holds when every member of $\mathbb{A}_N$ is doubly-stochastic, because such matrices always leave the uniform distribution invariant. (B^{++}) therefore holds for the ARPF, which has $\mathbb{A}_N = \{\mathrm{Id}, \mathbf{1}_{1/N}\}$.

To get some feel for why (B^{++}) is a natural condition for convergence, note that plugging the particle approximation $\pi_{n-1} \approx \pi^N_{n-1} = \sum_i W^i_{n-1}\delta_{\zeta^i_{n-1}} / \sum_i W^i_{n-1}$ into equation (2) for the predictor yields a finite mixture approximation of π_n

$$\pi_n \approx \frac{\sum_i W^i_{n-1} g_{n-1}(\zeta^i_{n-1}) f(\zeta_{n-1}, \cdot)}{\sum_i W^i_{n-1} g_{n-1}(\zeta^i_{n-1})}.$$

Under condition (B^{++}) the stochastic matrix $\alpha_{n-1} \in \mathbb{A}_N$ is doubly stochastic, hence

$$\frac{\sum_j W^j_{n-1} g^j_{n-1}(\zeta^j_{n-1}) f(\zeta^j_{n-1}, \cdot)}{\sum_j W^j_{n-1} g^j_{n-1}(\zeta^j_{n-1})} = \frac{\sum_i \sum_j \alpha^{ij}_{n-1} W^j_{n-1} g^j_{n-1}(\zeta^j_{n-1}) f(\zeta^j_{n-1}, \cdot)}{\sum_i \sum_j \alpha^{ij}_{n-1} W^j_{n-1} g^j_{n-1}(\zeta^j_{n-1})} = \frac{\sum_i W^i_n m^i_n}{\sum_i W^i_n},$$

where m^i_n is the distribution from which the new particle ζ^i_n is sampled, as in (16), justifying the particle approximation

$$\pi_n \approx \pi^N_n = \frac{\sum_i W^i_n \delta_{\zeta^i_n}}{\sum_i W^i_n}.$$

The main result of this section is:

Theorem 1. *Assume* (A2). *For any* $n \geq 0$, $\varphi \in \mathcal{L}$ *and* $r \geq 1$,

(1) *if* (B) *holds, then* $\mathbb{E}[Z^N_n] = Z_n$ *for any* $N \geq 1$,
(2) *if* (B^+) *holds, then*

$$\lim_{N\to\infty} \mathbb{E}\left[\left|Z^N_n - Z_n\right|^r\right] = 0, \tag{24}$$

$$\lim_{N\to\infty} \mathbb{E}\left[\left|\pi^N_n(\varphi) - \pi_n(\varphi)\right|^r\right] = 0, \tag{25}$$

and therefore $Z^N_n \to Z_n$ *and* $\pi^N_n(\varphi) \to \pi_n(\varphi)$ *in probability as* $N \to \infty$,

(3) *if* (B^{++}) *holds, then*

$$\sup_{N\geq 1}\sqrt{N}\mathbb{E}\big[\big|Z_n^N - Z_n\big|^r\big]^{1/r} < +\infty, \tag{26}$$

$$\sup_{N\geq 1}\sqrt{N}\mathbb{E}\big[\big|\pi_n^N(\varphi) - \pi_n(\varphi)\big|^r\big]^{1/r} < +\infty, \tag{27}$$

and therefore $Z_n^N \to Z_n$ *and* $\pi_n^N(\varphi) \to \pi_n(\varphi)$ *almost surely, as* $N \to \infty$.

Remark 1. The lack-of-bias property $\mathbb{E}[Z_n^N] = Z_n$ is desirable since it could be used to validate the use of αSMC within composite SMC/MCMC algorithms such as those of [1].

Remark 2. Theorem 1 holds without any sort of requirement that the entries of each α_n converge as $N \to \infty$. For example, (B^{++}) holds if for N odd we choose $\mathbb{A}_N = \{\mathrm{Id}\}$ and for N even we choose $\mathbb{A}_N = \{\mathbf{1}_{1/N}\}$. As a reflection of this, and as is apparent upon inspection of the proof, without further assumption we cannot in general replace $\sup_{N\geq 1}$ in (26)–(27) with $\lim_{N\to\infty}$, because such limits may not exist.

The following notation is used throughout the remainder of the paper. Introduce the non-negative kernels

$$Q_n : \mathsf{X}\times\mathcal{X} \to \mathbb{R}_+, \qquad Q_n\big(x, \mathrm{d}x'\big) := g_{n-1}(x) f\big(x, \mathrm{d}x'\big), \qquad n \geq 1, \tag{28}$$

the corresponding operators on functions and measures:

$$Q_n(\varphi)(x) := \int_{\mathsf{X}} Q_n\big(x, \mathrm{d}x'\big)\varphi\big(x'\big), \qquad \varphi \in \mathcal{L}, \tag{29}$$

$$\mu Q_n(\cdot) := \int_{\mathsf{X}} \mu(\mathrm{d}x) Q_n(x, \cdot), \qquad \mu \in \mathcal{M}, \tag{30}$$

and for $n \geq 1$ and $0 \leq p < n$,

$$Q_{p,p} := \mathrm{Id}, \qquad Q_{p,n} := Q_{p+1}\cdots Q_n. \tag{31}$$

We shall also consider the following scaled versions of these operators:

$$\overline{Q}_n := \frac{Q_n}{\pi_{n-1}(g_{n-1})}, \qquad \overline{Q}_{p,p} := \mathrm{Id}, \qquad \overline{Q}_{p,n} := \overline{Q}_{p+1}\cdots\overline{Q}_n. \tag{32}$$

Then define the non-negative measures

$$\gamma_n := \mu_0 Q_{0,n}(\cdot), \qquad n \geq 0,$$

under (A1) we are assured that $\gamma_n(1) > 0$. Due to the conditional independence structure of the HMM, it can easily be checked that

$$\pi_n = \frac{\gamma_n}{\gamma_n(1)}, \qquad Z_n = \gamma_n(1), \qquad n \geq 0$$

and

$$\overline{Q}_{p,n} = \frac{Q_{p,n}}{\pi_p Q_{p,n}(1)}.$$

For $i \in [N]$ and $0 \le p \le n$, introduce the random measures

$$\Gamma^N_{p,n} := \sum_i \beta^i_{p,n} W^i_p \delta_{\zeta^i_p}, \qquad \overline{\Gamma}^N_{p,n} := \frac{\Gamma^N_{p,n}}{\gamma_p(1)}, \tag{33}$$

where W^i_p is as in (5). For simplicity of notation, we shall write $\Gamma^N_n := \Gamma^N_{n,n}$, $\overline{\Gamma}^N_n := \overline{\Gamma}^N_{n,n}$. If we define

$$\overline{W}^i_n := \frac{W^i_n}{\gamma_n(1)}, \qquad n \ge 0, \tag{34}$$

then we have from (33),

$$\overline{\Gamma}^N_{p,n} = \sum_i \beta^i_{p,n} \overline{W}^i{}_p \delta_{\zeta^i_p}.$$

Finally, we observe from (21) that

$$\Gamma^N_n = \sum_i \beta^i_{n,n} W^i_n \delta_{\zeta^i_n} = N^{-1} \sum_i W^i_n \delta_{\zeta^i_n}.$$

Error decomposition

Throughout this section, let $\varphi \in \mathcal{L}$, $n \ge 0$ and $N \ge 1$ be arbitrarily chosen, but then fixed. Define, for $1 \le p \le n$ and $i \in [N]$,

$$\Delta^i_{p,n} := \overline{Q}_{p,n}(\varphi)\big(\zeta^i_p\big) - \frac{\sum_j \alpha^{ij}_{p-1} W^j_{p-1} \overline{Q}_{p-1,n}(\varphi)(\zeta^j_{p-1})}{\sum_j \alpha^{ij}_{p-1} W^j_{p-1} \overline{Q}_p(1)(\zeta^j_{p-1})},$$

and $\Delta^i_{0,n} := \overline{Q}_{0,n}(\varphi)(\zeta^i_0) - \mu_0 \overline{Q}_{0,n}(\varphi)$, so that $\mathbb{E}[\Delta^i_{p,n} | \mathcal{F}_{p-1}] = 0$ for any $i \in [N]$ and $0 \le p \le n$. Then for $0 \le p \le n$ and $i \in [N]$ set $k := pN + i$, and

$$\xi^N_k := \sqrt{N} \beta^i_{p,n} \overline{W}^i{}_p \Delta^i_{p,n},$$

so as to define a sequence $\{\xi^N_k; k = 1, \ldots, (n+1)N\}$. For $k = 1, \ldots, (n+1)N$, let $\mathcal{F}^{(k)}$ be the σ-algebra generated by $\{\zeta^i_p;\ pN + i \le k, i \in [N], 0 \le p \le n\}$. Set $\mathcal{F}^{(-1)} := \{\mathbb{X}, \varnothing\}$.

The following proposition is the main result underlying Theorem 1. The proof is given in the Appendix.

Proposition 1. *Assume* (A2) *and* (B). *We have the decomposition*

$$\sqrt{N}\big[\Gamma_n^N(\varphi) - \pi_n(\varphi)\big] = \sum_{k=1}^{(n+1)N} \xi_k^N, \tag{35}$$

where for $k = 1, \ldots, (n+1)N$, *the increment* ξ_k^N *is measurable w.r.t.* $\mathcal{F}^{(k)}$ *and satisfies*

$$\mathbb{E}\big[\xi_k^N | \mathcal{F}^{(k-1)}\big] = \mathbb{E}\big[\xi_k^N | \mathcal{F}_{p-1}\big] = 0 \qquad \textit{with } p := \big\lfloor (k-1)/N \big\rfloor. \tag{36}$$

For each $r \geq 1$ *there exists a universal constant* $B(r)$ *such that*

$$\begin{aligned} &\mathbb{E}\big[\big|\overline{\Gamma}_n^N(\varphi) - \pi_n(\varphi)\big|^r\big]^{1/r} \\ &\quad \leq B(r)^{1/r} \sum_{p=0}^{n} \operatorname{osc}\big(\overline{Q}_{p,n}(\varphi)\big) \mathbb{E}\left[\left|\sum_i \big(\beta_{p,n}^i \overline{W}_p^i\big)^2\right|^{r/2}\right]^{1/r}. \end{aligned} \tag{37}$$

The proof of Theorem 1, which is mostly technical, is given in the Appendix. Here we briefly discuss our assumptions and sketch some of the main arguments. Part (1) of Theorem 1 follows immediately from (35) and (36) applied with $\varphi = 1$. In turn, the martingale structure of (35) and (36) is underpinned by the measurability conditions (A2) and (B). The proofs of parts (2) and (3) of Theorem 1, involve applying Proposition 1 in conjunction with the identities

$$\begin{aligned} Z_n^N - Z_n &= \Gamma_n^N(1) - \gamma_n(1), \\ \pi_n^N(\varphi) - \pi_n(\varphi) &= \frac{\Gamma_n^N(\varphi)}{\Gamma_n^N(1)} - \frac{\gamma_n(\varphi)}{\gamma_n(1)}. \end{aligned} \tag{38}$$

In order to prove that these errors convergence to zero in probability, we show that the quadratic variation term in (37) converges to zero. In general, we cannot hope for the latter convergence without some sort of negligibility hypothesis on the product terms $\{\operatorname{osc}(\overline{Q}_{p,n}(\varphi))\beta_{p,n}^i \overline{W}_p^i; i \in [N]\}$. Assumption (A1) allows us to crudely upper-bound $\operatorname{osc}(\overline{Q}_{p,n}(\varphi))$ and $\overline{W}_p^i$; the measurability condition (B) allows us to dispose of the expectation in (37); then via Markov's inequality and the classical equivalence:

$$\lim_{N\to\infty} \max_{i\in[N]} \beta_{p,n}^i = 0 \quad \Leftrightarrow \quad \lim_{N\to\infty} \sum_i \big(\beta_{p,n}^i\big)^2 = 0,$$

which holds since $(\max_{i\in[N]} \beta_{p,n}^i)^2 \leq \sum_i (\beta_{p,n}^i)^2 \leq \max_{i\in[N]} \beta_{p,n}^i$, the negligibility part of (B^+) guarantees that $|\Gamma_n^N(\varphi) - \gamma_n(\varphi)|$ converges to zero in probability. The stronger condition (B^{++}) buys us the $\sqrt{N}$ scaling displayed in part (3). In Section 5.2, we discuss what can go wrong when (B^+) does not hold.

4. Stability

In this section, we study the stability of approximation errors under the following regularity condition.

Assumption. (C) *There exists* $(\delta, \epsilon) \in [1, \infty)^2$ *such that*

$$\sup_{n \geq 0} \sup_{x,y} \frac{g_n(x)}{g_n(y)} \leq \delta, \qquad f(x, \cdot) \leq \epsilon f(y, \cdot), \qquad (x, y) \in \mathsf{X}^2.$$

(C) is a standard hypothesis in studies of non-asymptotic stability properties of SMC algorithms. Similar conditions have been adopted in [8], Chapter 7, and [18], amongst others. (C) guarantees that $Q_{p,n}$, and related objects, obey a variety of regularity conditions. In particular, we immediately obtain

$$\sup_{p,n} \sup_{x} \overline{Q}_{p,n}(1)(x) \leq \sup_{p,n} \sup_{x,y} \frac{Q_{p,n}(1)(x)}{Q_{p,n}(1)(y)} \leq \delta\epsilon < +\infty. \tag{39}$$

Furthermore, if we introduce the following operators on probability measures:

$$\Phi_n : \mu \in \mathcal{P} \mapsto \frac{\mu Q_n}{\mu(g_{n-1})} \in \mathcal{P}, \qquad n \geq 1, \tag{40}$$

$$\Phi_{p,n} := \Phi_n \circ \cdots \circ \Phi_{p+1}, \qquad 0 \leq p < n. \tag{41}$$

It is well-known that under (C), $\Phi_{p,n}$ is uniformly exponentially stable, in the sense of the somewhat crude estimate in the following lemma.

Lemma 2. *Assume* (C). *Then there exists a finite constant* C *and* $\rho \in [0, 1)$ *such that*

$$\sup_{\mu, \mu' \in \mathcal{P}} \left\| \Phi_{p,n}(\mu) - \Phi_{p,n}\left(\mu'\right) \right\|_{\mathrm{tv}} \leq C\rho^{n-p}.$$

For a proof see, for example, [8], Proposition 4.3.6. It follows from (2), (40) and (41) that

$$\pi_{n+1} = \Phi_{n+1}(\pi_n) = \Phi_{p,n+1}(\pi_p) = \Phi_{0,n+1}(\mu_0), \qquad 0 \leq p \leq n,$$

so Lemma 2 can be used to describe the forgetting of the initial distribution of the non-linear filter. Properties similar to (39) and the exponential stability in Lemma 2 can be obtained under conditions weaker and more realistic than (C), see, for example, [22] but the developments involved are substantially more technical, lengthy and complicated to present. Our aim is to expedite the presentation of stability properties of αSMC, and (C) allows this to be achieved while retaining some of the essence of more realistic hypotheses on g_n and f.

The main result of this section is the following theorem, whose proof we briefly postpone.

Theorem 2. *Assume* (A2), (B^{++}) *and* (C). *Then there exist finite constants*, c_1 *and for each* $r \geq 1$, $c_2(r)$, *such that for any* $\tau \in (0, 1]$, $N \geq 1$, *and* $\varphi \in \mathcal{L}$,

$$\inf_{n\geq 0} \mathcal{E}_n^N \geq \tau \quad \Rightarrow \quad \begin{cases} \sup_{n\geq 1} \mathbb{E}\left[\left(\frac{Z_n^N}{Z_n}\right)^2\right]^{1/n} \leq 1 + \frac{c_1}{N\tau} \quad \text{and} \\ \sup_{n\geq 0} \mathbb{E}\left[\left|\pi_n^N(\varphi) - \pi_n(\varphi)\right|^r\right]^{1/r} \leq \|\varphi\| \frac{c_2(r)}{\sqrt{N\tau}}. \end{cases} \tag{42}$$

Remark 3. It follows quite immediately from the first inequality of (42) that

$$\left.\begin{array}{r} \inf_{n\geq 0} \mathcal{E}_n^N \geq \tau \\ \text{and} \\ N\tau \geq nc_1 \end{array}\right\} \quad \Rightarrow \quad \mathbb{E}\left[\left(\frac{Z_n^N}{Z_n} - 1\right)^2\right] \leq \frac{2nc_1}{N\tau},$$

see Lemma 6 in the Appendix.

Remark 4. It follows immediately from the second inequality in (42) that when $\inf_{n\geq 0} \mathcal{E}_n^N \geq \tau$ for all $N \geq 1$, the prediction filter errors are time-uniformly convergent in the sense

$$\lim_{N\to\infty} \sup_{n\geq 0} \mathbb{E}\left[\left|\pi_n^N(\varphi) - \pi_n(\varphi)\right|^r\right]^{1/r} = 0.$$

Remark 5. Further to the discussion of Section 2.2, in the case of the BPF we have $\mathcal{E}_n^N = 1$ and hence $\inf_{n\geq 0} \mathcal{E}_n^N \geq \tau$ always, and for the ARPF we also have $\inf_{n\geq 0} \mathcal{E}_n^N \geq \tau$ always, by virtue of the ESS rule for selection of α_n. In Section 5, we shall introduce new algorithms designed to guarantee $\inf_{n\geq 0} \mathcal{E}_n^N \geq \tau$.

Remark 6. It is possible to deduce estimates for the constants c_1 and $c_2(r)$ using the statements and proofs of Propositions 2 and 3, which are the main ingredients in the proof of Theorem 2. We omit such expressions only for simplicity of presentation.

The proofs of Propositions 2 and 3 are given in the Appendix.

Proposition 2. *Assume* (A2), (B^{++}) *and* (C). *If for some sequence of constants* $\{\tau_n; n \geq 0\} \in (0, 1]^{\mathbb{N}}$ *and* $N \geq 1$,

$$\mathcal{E}_n^N \geq \tau_n,$$

then for any $n \geq 1$,

$$\mathbb{E}\left[\left(\frac{Z_n^N}{Z_n} - 1\right)^2\right] \leq \sum_{p=0}^{n-1} \frac{\operatorname{osc}(\overline{Q}_{p,n}(1))^2}{N\tau_p} \left(\mathbb{E}\left[\left(\frac{Z_p^N}{Z_p} - 1\right)^2\right] + 1\right).$$

Proposition 3. *Consider the constants and Markov kernels*:

$$\delta_{p,n} := \sup_{x,y} \frac{Q_{p,n}(1)(x)}{Q_{p,n}(1)(y)}, \qquad P_{p,n}(x, A) := \frac{Q_{p,n}(\mathbb{I}_A)(x)}{Q_{p,n}(1)(x)}, \qquad x \in \mathsf{X}, A \in \mathcal{X}, 0 \le p \le n.$$

Assume (A2), (B) *and* (C). *Then for any* $r \ge 1$ *there exists a finite constant* $B(r)$ *such that for any* $N \ge 1$, $n \ge 0$, *and* $\varphi \in \mathcal{L}$,

$$\mathbb{E}\left[\left|\pi_n^N(\varphi) - \pi_n(\varphi)\right|^r\right]^{1/r} \le 4B(r)^{1/r} \sum_{p=0}^{n} \delta_{p,n} \left\| P_{p,n}(\bar{\varphi}) \right\| \mathbb{E}\left[\left|\mathcal{C}_{p,n}^N\right|^r\right]^{1/r}. \tag{43}$$

where $\bar{\varphi} := \varphi - \pi_n(\varphi)$ *and*

$$\mathcal{C}_{p,n}^N := \frac{\sqrt{\sum_i (\beta_{p,n}^i W_p^i)^2}}{\sum_i \beta_{p,n}^i W_p^i}.$$

Proof of Theorem 2. For the first bound on the right of (42) under the conditions of the theorem, we apply Proposition 2 to give the following recursive bound:

$$v_n \le \sum_{p=0}^{n-1} \frac{C}{N\tau}(v_p + 1), \tag{44}$$

where $v_n := \mathbb{E}[(Z_n^N/Z_n - 1)^2]$ and

$$C := \sup_{p,n} \operatorname{osc}\left(\overline{Q}_{p,n}(1)\right)^2 \le 4 \sup_{p,n} \left\| \overline{Q}_{p,n}(1) \right\|^2 < +\infty,$$

under (C); see (39). We shall now prove

$$v_n \le \left(1 + \frac{C}{N\tau}\right)^n - 1 \qquad \forall n \ge 0, \tag{45}$$

which holds trivially if $C = 0$, since in that case $v_n = 0$ by (44). Therefore suppose $C > 0$. The argument is inductive. To initialize, note that since by definition $Z_0^N = Z_0 = 1$, we have $v_0 = 0$. Now assume (45) holds at all ranks strictly less than some fixed $n \ge 1$. Using (44), we then have at rank n,

$$\begin{aligned} v_n &\le \frac{C}{N\tau} \sum_{p=0}^{n-1} (v_p + 1) \le \frac{C}{N\tau} \sum_{p=0}^{n-1} \left(1 + \frac{C}{N\tau}\right)^p \\ &= \frac{C}{N\tau} \frac{(1 + C/N\tau)^n - 1}{(1 + C/N\tau) - 1} \\ &= \left(1 + \frac{C}{N\tau}\right)^n - 1. \end{aligned}$$

This completes the proof of (45), from which the first inequality on the right of (42) follows immediately upon noting that by Theorem 1, $\mathbb{E}[Z_n^N] = Z_n$.

For the second bound on the right of (42), first note that by Lemma 2, under (C) we have

$$\begin{aligned}\left\|P_{p,n}(\bar{\varphi})\right\| &= \sup_x \left|P_{p,n}(\varphi)(x) - \pi_n(\varphi)\right| \\ &= \sup_x \left|\Phi_{p,n}(\delta_x)(\varphi) - \Phi_{p,n}(\pi_p)(\varphi)\right| \\ &\leq \sup_{\mu,\nu\in\mathcal{P}} \left\|\Phi_{p,n}(\mu) - \Phi_{p,n}(\nu)\right\|_{\mathrm{tv}} \|\varphi\| \leq \|\varphi\| C\rho^{n-p},\end{aligned}$$

and by (39),

$$\sup_{n\geq 0}\sup_{p\leq n} \delta_{p,n} < +\infty.$$

Using these upper bounds, the fact that under (B^{++}) we have $\beta_{p,n}^i = 1/N$, and Proposition 3, we find that there exists a finite constant $\widetilde{B}(r)$ such that for any $N \geq 1$, $n \geq 0$, $\varphi \in \mathcal{L}$,

$$\mathbb{E}\left[\left|\pi_n^N(\varphi) - \pi_n(\varphi)\right|^r\right]^{1/r} \leq \|\varphi\| \frac{\widetilde{B}(r)}{\sqrt{N}} \sum_{p=0}^{n} \rho^{n-p} \mathbb{E}\left[\left|\mathcal{E}_p^N\right|^{-r/2}\right]^{1/r},$$

where

$$\mathcal{E}_n^N = \frac{(N^{-1}\sum_i W_n^i)^2}{N^{-1}\sum_i (W_n^i)^2}.$$

□

5. Discussion

5.1. Why not just run independent particle filters and average?

One obvious approach to parallelization of SMC is to run a number of independent copies of a standard algorithm, such as the BPF, and then in some sense simply average their outputs. Let us explain possible shortcomings of this approach.

Suppose we want to run $s \geq 1$ independent copies of Algorithm 3, each with $q \geq 1$ particles. For purposes of exposition, it is helpful to express this collection of independent algorithms as a particular instance of αSMC: for the remainder of Section 5.1, we set $N = sq$ and consider Algorithm 1 with $\mathbb{A}_N$ chosen to consist only of the block diagonal matrix:

$$\begin{bmatrix} \mathbf{q^{-1}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{q^{-1}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{q^{-1}} \end{bmatrix}, \tag{46}$$

where $\mathbf{q}^{-1}$ is a $q \times q$ submatrix with every entry equal to q^{-1} and $\mathbf{0}$ is a submatrix of zeros, of the same size. In this situation, a simple application of Lemma 1 shows that for any $n \geq 1$ and $\ell \in [s]$, if we define $B(\ell) := \{(\ell-1)q+1, (\ell-1)q+2, \ldots, \ell q\}$, then

$$\text{for all } i_n \in B(\ell), \qquad W_n^{i_n} = \prod_{p=0}^{n-1} \left(N^{-1} \sum_{i_p \in B(\ell)} g_p\big(\zeta_p^{i_p}\big) \right) =: \mathbb{W}_n^{\ell}, \tag{47}$$

cf. (11)–(12), and furthermore upon inspection of Algorithm 1, we find

$$\text{for all } \ell \in [s] \text{ and } i \in B(\ell) \qquad \mathbb{P}\big(\zeta_n^i \in A | \mathcal{F}_{n-1}\big) = \frac{\sum_{j \in B(\ell)} g_{n-1}(\zeta_{n-1}^j) f(\zeta_{n-1}^j, A)}{\sum_{j \in B(\ell)} g_{n-1}(\zeta_{n-1}^j)}, \tag{48}$$

for any $A \in \mathcal{X}$. It follows that the blocks of particles

$$\hat{\zeta}_n^k := \big\{\zeta_n^i\big\}_{i \in B(\ell)}, \qquad \ell \in [s],$$

are independent, and for each $\ell \in [s]$, the sequence $\{\hat{\zeta}_n^{\ell}; n \geq 0\}$ evolves under the same law as a BPF, with q particles. Furthermore, we notice

$$\pi_n^N = \pi_n^{sq} = \frac{\sum_i W_n^i \delta_{\zeta_n^i}}{\sum_i W_n^i} = \frac{\sum_{\ell \in [s]} \sum_{i \in B(\ell)} W_n^i \delta_{\zeta_n^i}}{\sum_{\ell \in [s]} \sum_{i \in B(\ell)} W_n^i} = \frac{\sum_{\ell \in [s]} \mathbb{W}_n^{\ell} (q^{-1} \sum_{i \in B(\ell)} \delta_{\zeta_n^i})}{\sum_{\ell \in [s]} \mathbb{W}_n^{\ell}},$$

where $q^{-1} \sum_{i \in B(\ell)} \delta_{\zeta_n^i}$ may be regarded as the approximation of π_n obtained from the ℓth block of particles. Since we have assumed that $\mathbb{A}_N$ consists only of the matrix (46), (A2) and (B^{++}) hold, and by Theorem 1 we are assured of the a.s. convergence $\pi_n^{sq}(\varphi) \to \pi_n(\varphi)$ when q is fixed and $s \to \infty$. In words, we have convergence as the total number of bootstrap algorithms tends to infinity, even though the number of particles within each algorithm is fixed. On the other hand, simple averaging of the output from the s independent algorithms would entail reporting:

$$\frac{1}{sq} \sum_{i \in [sq]} \delta_{\zeta_n^i} \tag{49}$$

as an approximation of π_n; the problem is that (49) is biased, in the sense that in general it is not true that, with q fixed, $(sq)^{-1} \sum_{i \in [sq]} \varphi(\zeta_n^i) \to \pi_n(\varphi)$ as $s \to \infty$ (although obviously we do have convergence if $q \to \infty$). In summary, simple averages across independent particle filters do not, in general, converge as the number of algorithms grows.

We can also discuss the quality of an approximation of Z_n obtained by simple averaging across the s independent algorithms; let us consider the quantities

$$\mathbb{Z}_n^{(q,\ell)} := \frac{1}{\ell} \sum_{j \in [\ell]} \mathbb{W}_n^j, \qquad \ell \in [s].$$

Comparing (47) with (12), and noting (48) and the independence properties described above, we have

$$\mathbb{E}\big[\mathbb{Z}_n^{(q,s)}\big] = Z_n, \qquad \mathbb{E}\left[\left(\frac{\mathbb{Z}_n^{(q,s)}}{Z_n} - 1\right)^2\right] = \frac{1}{s}\mathbb{E}\left[\left(\frac{\mathbb{Z}_n^{(q,1)}}{Z_n} - 1\right)^2\right], \tag{50}$$

where the first equality holds due to the first part of Theorem 1: in this context the well known lack-of-bias property of the BPF. Under certain ergodicity and regularity conditions $\mathbb{E}[(\mathbb{Z}_n^{(q,1)}/Z_n)^2]$ can grow exponentially fast along observation sample paths when q is fixed and $n \to \infty$ [23]. When that occurs, it is clear from (49) that s must be scaled up exponentially fast with n in order to control the relative variance of $\mathbb{Z}_n^{(q,s)}$. On the other hand, by Theorem 2 and Remark 3, it is apparent that if we design an instance of αSMC so as to enforce $\inf_{n\geq 0} \mathcal{E}_n^N > 0$, then we can control $\mathbb{E}[(Z_n^N/Z_n)^2]$ at a more modest computational cost. When $\mathbb{A}_N$ consists only of the matrix (46) we do not have a guarantee that $\inf_{n\geq 0} \mathcal{E}_n^N > 0$, but in Section 5.3 we shall suggest some novel algorithms which do guarantee this lower bound and therefore enjoy the time-uniform convergence and linear-in-time variance properties of Theorem 2. Before addressing these stability issues, we discuss the conditions under which the αSMC algorithm converges.

5.2. Ensuring convergence

Throughout Section 5.2, we consider the generic Algorithm 1. We describe what can go wrong if the conditions (B^+) and (B^{++}) of Theorem 1 do not hold: suppose that $\mathbb{A}_N$ consists only of the transition matrix of a simple random walk on the star graph with N vertices, call it $\mathcal{S}_N$. That is, for $N > 2$, $\mathcal{S}_N$ is an undirected tree with one internal vertex and $N - 1$ leaves, and for $N \leq 2$, all vertices are leaves. Examples of $\mathcal{S}_N$ are illustrated in Figure 1. It is elementary that a simple random walk on $\mathcal{S}_N$ has unique invariant distribution given by

$$\frac{d_N^i}{\sum_j d_N^j}, \qquad i \in [N], \text{ where } d_N^i := \text{ degree of vertex } i \text{ in } \mathcal{S}_N,$$

so that (B^{++}) does not hold for $N > 2$. Assuming that for every $N > 2$, the internal vertex of $\mathcal{S}_N$ is labelled vertex 1, then elementary calculations show that

$$\max_{i\in[N]} \beta_{p,n}^i = \beta_{p,n}^1 = \begin{cases} \dfrac{N-1}{N}, & \text{if } (n-p) \text{ is even,} \\ \dfrac{1}{N}, & \text{if } (n-p) \text{ is odd,} \end{cases}$$

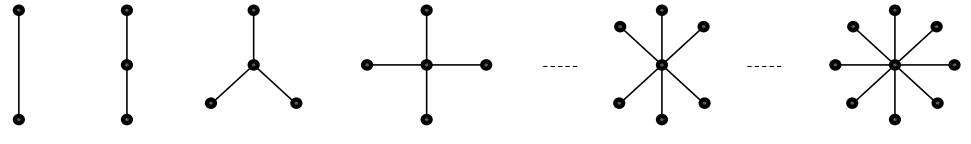

Figure 1. Star graphs.

so (B^+) also does not hold, and thus part (2) of Theorem 1 does not hold.

As a more explicit example of convergence failure, suppose that $\mathbb{A}_N$ consists only of the matrix which has 1 for every entry in its first column, and zeros for all other entries. This is the transition matrix of a random walk on a directed graph of which all edges lead to vertex 1. It follows that for all $0 \leq p < n$, we have $\beta^1_{p,n} = 1$ and $\beta^i_{p,n} = 0$ for all $i \in [N] \setminus \{1\}$, so ($B^+$) clearly does not hold. If additionally $f(x, \cdot) = \delta_x(\cdot)$, then by inspection of Algorithm 1 we have $\mathbb{P}(\{\zeta^i_n = \zeta^1_0\}) = 1$ for all $i \in [N]$ and all $n \geq 1$. We then also have $\mathbb{P}(\{\pi^N_n = \delta_{\zeta^1_0}\}) = 1$, so that we obtain a generally poor and non-convergent approximation of π_n.

In both these situations vertex 1 is, in graph theoretic terms, a *hub* and an intuitive explanation of the convergence failure is that the contribution of particle 1 to π^N_n does not become negligible as $N \to \infty$, so that no "averaging" takes place. Assumption (B^+) ensures enough negligibility to prove the weak laws of large numbers in Theorem 1. Assumption (B^{++}) may be viewed as ensuring negligibility, and in such a way as to ensure the $\sqrt{N}$ rate of convergence and strong law in the final part of Theorem 1. As a practical summary, we recommend verifying (B^{++}), or at least avoid using graphs with hubs, since otherwise αSMC may not converge.

5.3. Provably stable algorithms with adaptive interaction

There are of course many choices of $\mathbb{A}_N$ which do satisfy (B^{++}). In this section, we provide some guidance and suggestions on this matter. In order to focus our attention, we consider in addition to (B^{++}), the following criteria against which to assess candidates for $\mathbb{A}_N$ and whatever functional is used at line ($\star$) of Algorithm 1:

(a) the condition $\inf_{n \geq 0} \mathcal{E}^N_n > 0$ should be enforced, so as to ensure stability,
(b) the computational complexity of associated sampling, weight and ESS calculations should not be prohibitively high.

The motivation for (a) is the theoretical assurance given by Theorem 2. The motivation for (b) is simply that we do not want an algorithm which is much more expensive than any of the standard SMC methods, Algorithms 2–3 and the ARPF. It is easily checked that the complexity of SIS is $O(N)$ per unit time step, which is the same as the complexity of the BPF [4] and the ARPF.

Throughout the remainder of Section 5.3, we shall assume that $\mathbb{A}_N$ consists only of transition matrices of simple random walks on regular undirected graphs. We impose a little structure in addition to this as per the following definition, which identifies an object related to the standard notion of a block-diagonal matrix.

Definition. *A* B-matrix *is a Markov transition matrix which specifies a simple random walk on a regular undirected graph which has a self-loop at every vertex and whose connected components are all complete subgraphs.*

Note that due to the graph regularity appearing in this definition, if $\mathbb{A}_N$ consists only of B-matrices, then (B^{++}) is immediately satisfied. This regularity is also convenient for purposes of interpretation: it seems natural to use graph degree to give a precise meaning to "degree of interaction". Indeed Id and $\mathbf{1}_{1/N}$ are both B-matrices, respectively, specifying simple random

walks on 1-regular and N-regular graphs, and recall for the ARPF, $\mathbb{A}_N = \{\mathrm{Id}, \mathbf{1}_{1/N}\}$; the main idea behind the new algorithms below is to consider an instance of αSMC in which $\mathbb{A}_N$ is defined to consist of B-matrices of various degrees $d \in [N]$, and define adaptive algorithms which select the value of α_{n-1} by searching through $\mathbb{A}_N$ to find the graph with the smallest d which achieves $\mathcal{E}_n^N \geq \tau > 0$ and hence satisfies criterion (a). In this way, we ensure provable stability whilst trying to avoid the complete interaction which occurs when $\alpha_{n-1} = \mathbf{1}_{1/N}$.

Another appealing property of B-matrices is formalized in the following lemma; see criterion (b) above. The proof is given in the Appendix.

Lemma 3. *Suppose that $A = (A^{ij})$ is a B-matrix of size N. Then given the quantities $\{W_{n-1}^i\}_{i\in[N]}$ and $\{g_{n-1}(\zeta_{n-1}^i)\}_{i\in[N]}$, the computational complexity of calculating $\{W_n^i\}_{i\in[N]}$ and simulating $\{\zeta_n^i\}_{i\in[N]}$ as per Algorithm 1, using $\alpha_{n-1} = A$, is $\mathrm{O}(N)$.*

When calculating the overall complexity of Algorithm 1 we must also consider the complexity of line ($\star$), which in general depends on $\mathbb{A}_N$ and the particular functional used to choose α_n. We resume this complexity discussion after describing the specifics of some adaptive algorithms.

Adaptive interaction.

Throughout this section, we set $m \in \mathbb{N}$ and then $N = 2^m$. Consider Algorithm 1 with $\mathbb{A}_N$ chosen to be the set of B-matrices of size N. We suggest three adaptation rules at line ($\star$) of Algorithm 1: Simple, Random, and Greedy, all implemented via Algorithm 4 (note that dependence of some quantities on n is suppressed from the notation there), but differing in the way they select the index list $\mathcal{I}_k$ which appears in the "while" loop of that procedure. The methods for selecting $\mathcal{I}_k$ are summarised in Table 1: the Simple rule needs little explanation, the Random rule implements an independent random shuffling of indices and the Greedy rule is intended, heuristically, to pair large weights, $\mathbb{W}_k^i$, with small weights in order to terminate the "while" loop with as small a value of k as possible. Note that, formally, in order for our results for αSMC to apply when the Random rule is used, the underlying probability space must be appropriately extended, but the details are trivial so we omit them.

Following the termination of the "while" loop, Algorithm 4 outputs an integer K_{n-1} and a partition $\{B(K_{n-1}, i)\}_{i\in[N/2^{K_{n-1}}]}$ of $[N]$ into $N/2^{K_{n-1}}$ subsets, each of cardinality $2^{K_{n-1}}$; this partition specifies α_{n-1} as a B-matrix and $2^{K_{n-1}}$ is the degree of the corresponding graph (we keep track of K_{n-1} for purposes of monitoring algorithm performance in Section 5.4). Proposition 4 is a formal statement of its operation and completes our complexity considerations. The proof is given in the Appendix. It can be checked by an inductive argument similar to the proof of Lemma 4, also in the Appendix, that when α_n is chosen according to Algorithm 4 combined with any of the adaptation rules in Table 1, (A2) is satisfied.

Proposition 4. *The weights $\{\mathbb{W}_k^i\}_{i\in[N/2^k]}$ calculated in Algorithm 4 obey the expression*

$$\mathbb{W}_k^i = 2^{-k} \sum_{j\in B(k,i)} W_{n-1}^j g_{n-1}\big(\zeta_{n-1}^j\big). \tag{51}$$

Algorithm 4 Adaptive selection of α_{n-1}

At iteration n and line $(\star)$ of Algorithm 1
  For $i = 1, \ldots, N$,
    Set $B(0, i) = \{i\}$, $\mathbb{W}_0^i = W_{n-1}^i g_{n-1}(\zeta_{n-1}^i)$,
  Set $k = 0$,
  Set $\overline{\mathbb{W}}_0 = N^{-1} \sum_i \mathbb{W}_0^i$, $\mathcal{E} = \frac{(\overline{\mathbb{W}}_0)^2}{N^{-1} \sum_i (\mathbb{W}_0^i)^2}$,
  While $\mathcal{E} < \tau$
    Set $\mathcal{I}_k$ according to the Simple, Random or Greedy scheme of Table 1
    For $i = 1, \ldots, N/2^{k+1}$
      Set $B(k+1, i) = B(k, \mathcal{I}_k(2i-1)) \cup B(k, \mathcal{I}_k(2i))$
      Set $\mathbb{W}_{k+1}^i = \mathbb{W}_k^{\mathcal{I}_k(2i-1)}/2 + \mathbb{W}_k^{\mathcal{I}_k(2i)}/2$
    Set $k = k + 1$
    Set $\mathcal{E} = \frac{(\overline{\mathbb{W}}_0)^2}{N^{-1} 2^k \sum_{i \in [N/2^k]} (\mathbb{W}_k^i)^2}$
  Set $K_{n-1} = k$
  Set $\alpha_{n-1}^{ij} = \begin{cases} 1/2^{K_{n-1}}, & \text{if } i \sim j \text{ according to } \{B(K_{n-1}, i)\}_{i \in [N/2^{K_{n-1}}]}, \\ 0, & \text{otherwise.} \end{cases}$

Moreover, α_{n-1} delivered by Algorithm 4 is a B-matrix and when this procedure is used at line $(\star)$ of Algorithm 1, the weights calculated in Algorithm 1 are given, for any $i \in [N/2^{K_{n-1}}]$, by

$$W_n^j = \mathbb{W}_{K_{n-1}}^i \qquad \text{for all } j \in B(K_{n-1}, i) \tag{52}$$

and $\mathcal{E}_n^N \geq \tau$ always. The overall worst-case complexity of Algorithm 1 is, for the three adaptation rules in Table 1, Simple: $\mathrm{O}(N)$, *Random*: $\mathrm{O}(N)$, *and Greedy*: $\mathrm{O}(N \log_2 N)$.

5.4. Numerical illustrations

We consider a stochastic volatility HMM:

$$X_0 \sim \mathcal{N}(0, 1), \qquad X_n = a X_{n-1} + \sigma V_n,$$
$$Y_n = \varepsilon W_n \exp(X_n/2),$$

Table 1. Adaptation rules for choosing $\mathcal{I}_k$

Simple	set $\mathcal{I}_k = (1, \ldots, N/2^k)$
Random	if $k = 0$, set $\mathcal{I}_k$ to a random permutation of $[N/2^k]$, otherwise $\mathcal{I}_k = (1, \ldots, , N/2^k)$
Greedy	set $\mathcal{I}_k$ such that $\mathbb{W}_k^{\mathcal{I}_k(1)} \geq \mathbb{W}_k^{\mathcal{I}_k(3)} \geq \cdots \geq \mathbb{W}_k^{\mathcal{I}_k(N/2^k-1)} \geq \mathbb{W}_k^{\mathcal{I}_k(N/2^k)} \geq \cdots \geq \mathbb{W}_k^{\mathcal{I}_k(4)} \geq \mathbb{W}_k^{\mathcal{I}_k(2)}$

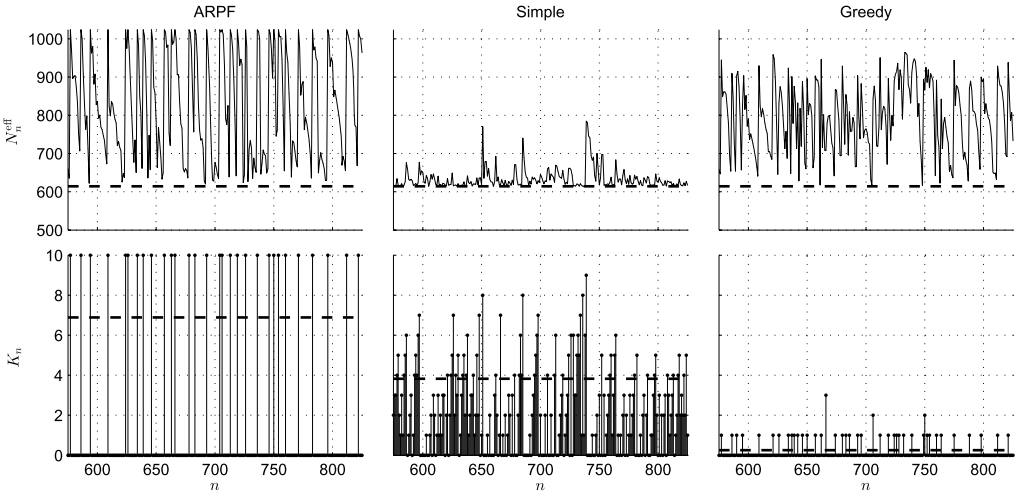

Figure 2. Snapshots of ESS and degree of interaction. Top: N_n^{eff} vs. n (solid) and threshold τN (dashed). Bottom: K_n vs. n (stems) and the base two logarithm of the time-average of 2^{K_n} (dashed). Recall from Section 5.3 that 2^{K_n} is the degree of the graph corresponding to the matrix α_n selected by Algorithm 4, and returned to line ($\star$) of Algorithm 1.

where $\{V_n\}_{n\in\mathbb{N}}$ and $\{W_n\}_{n\in\mathbb{N}}$ are sequences of mutually i.i.d. $\mathcal{N}(0, 1)$ random variables, $|a| < 1$, and $\sigma, \varepsilon > 0$. To study the behaviour of the different adaptation rules in terms of effective sample size, a sequence of $3 \cdot 10^4$ observations were generated from the model with $a = 0.9$, $\sigma = 0.25$, and $\varepsilon = 0.1$. This model obviously does not satisfy (C), but (A1) is satisfied as long as the observation record does not include the value zero.

The ARPF and αSMC with the Simple, Random and Greedy adaptation procedures specified in Section 5.3 were run on this data with $N = 2^{10}$ and threshold $\tau = 0.6$. To give some impression of ESS and interaction behaviour, Figure 2 shows snapshots of N_n^{eff} and K_n versus n, for $575 \leq n \leq 825$. The sample path of N_n^{eff} for ARPF displays a familiar saw-tooth pattern, jumping back up to $N = 2^{10}$ when resampling, that is, when $K_n = 10$. The Simple adaptation scheme keeps N_n^{eff} just above the threshold $\tau N = 0.6 \times 2^{10}$, whereas the Greedy strategy is often able to keep N_n^{eff} well above this threshold, with smaller values of K_n, that is, with a lower degree of interaction. The results for the Random adaptation rule, not shown in this plot, where qualitatively similar to those of the Greedy algorithm but slightly closer to the Simple adaptation.

In order to examine the stationarity of the particle processes as well as the statistical behavior of the degree of interaction over time, Figure 3 shows two histograms of K_n for each of the adaptation rules. One histogram is based on the sample of K_n where $100 < n \leq 15\,050$, and the other is based on K_n where $15\,050 < n \leq 30\,000$. For each algorithm, the similarity between the histograms for the two time intervals suggests that the process $\{K_n\}_{n\geq 0}$ is stationary. As expected, the distribution of K_n for ARPF is dichotomous taking only values equal to $K_n = 0$ when there is no interaction, that is, the resampling is skipped or $K_n = 10$ for the complete interaction, that is, resampling. It is apparent that the Simple, Random and Greedy algorithms move the distribution of K_n toward smaller values and almost always manage to avoid the complete interaction. For

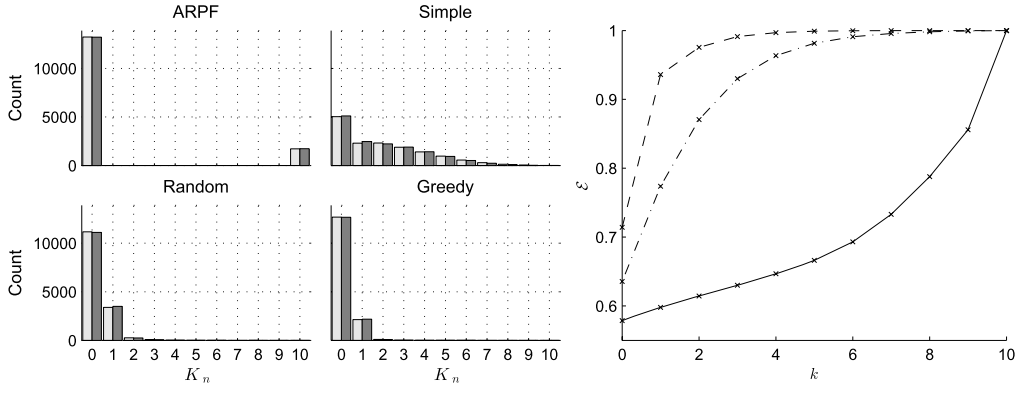

Figure 3. Left: Histograms of K_n for the ARPF and the three adaptation rules of Table 1. The light bars were obtained from $\{K_n; n = 101, \ldots, 15\,050\}$ and the dark bars from $\{K_n; n = 15\,051, \ldots, 30\,000\}$ Right: Growth of $\mathcal{E}$ vs. k for the Simple (solid), Random (dash-dot) and Greedy (dashed).

the Random and Greedy algorithms, K_n rarely exceeds 1, that is, in order to guarantee $\mathcal{E}_n^N$ it is rarely necessary to consider anything more than pair-wise interaction.

The plot on the right of Figure 3 shows, for each of the Simple, Random and Greedy adaptation rules, the relationship between the intermediate variables $\mathcal{E}$ and k appearing in the "while" loop of Algorithm 4. In order to obtain equal sample sizes for plotting purposes, Algorithm 4 was modified slightly so as to evaluate $\mathcal{E}$ for every value $k \in \{0, \ldots, m\}$, whilst still outputting K_{n-1} as the smallest value of k achieving $\mathcal{E} \geq \tau$. The plotted data were then obtained, for each k, by averaging the corresponding values of $\mathcal{E}$ over the time steps of the algorithm. It is apparent that, for small values of k, the Random and Greedy strategies achieve a faster increase in $\mathcal{E}$ than the Simple strategy, and this explains the shape of the histograms on the left of Figure 3.

Figure 4 shows a comparison of the mean squared errors (MSE) of approximating the conditional expectation of $\phi(X_p)$ with respect to the underlying stochastic volatility HMM given the observations $\{y_n; 0 \leq n \leq p + \ell\}$, where $\ell \in \{-5, 0, 1\}$ and ϕ is some test function. The cases, $\ell = -5$, $\ell = 0$, and $\ell = 1$ correspond to the lag 5 smoother, filter and one step predictor, respectively. The lag 5 smoother results were obtained by tracing back ancestral lineages. In order to estimate the approximation error, a reference value for the conditional expectation was evaluated by running a BPF with a large sample size $N = 2^{17}$. Approximation errors were evaluated for $N_{\mathrm{MC}} = 1000$ Monte Carlo runs of 1000 time steps each with $N = 2^9$, and MSE was obtained by averaging over the time steps and the Monte Carlo runs. First 30 time steps were excluded in the calculations to avoid any non-stationary effects due to initialization. The results show that the Random and Greedy algorithms produce consistently smaller errors than the Simple algorithm and for large values of τ the Greedy algorithm appears to consistently outperform ARPF.

5.5. Concluding remarks

- The martingale decomposition presented in Proposition 1 may also be exploited to pursue central limit theorems. A study of this will be conducted elsewhere, but we believe, further

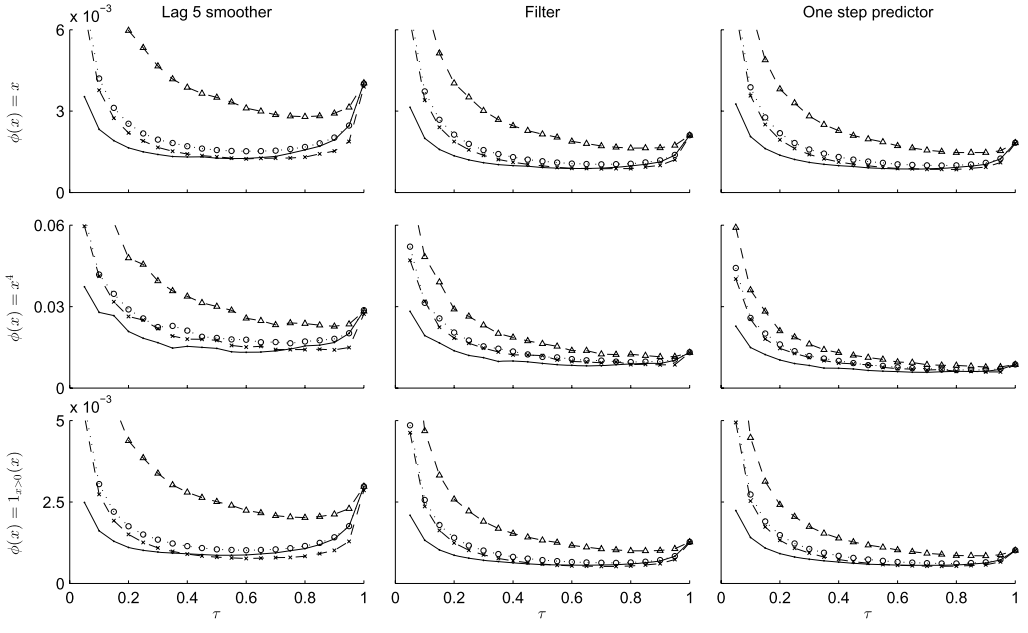

Figure 4. MSE vs. τ for the lag 5 smoother, filter, and one step predictor using the four algorithms ARPF (solid), Simple ($\triangle$), Random ($\circ$), and Greedy ($\times$) and three test functions ϕ.

to Remark 2, that it will in general involve some further hypotheses in order to ensure convergence of the covariance of this martingale and thus prove the existence of a well-defined asymptotic variance.

- It is worth pointing out that there are also SMC algorithms other than those listed in Section 2.2 that can be formulated as instances of αSMC, for example, the stratified resampling algorithm of Kitagawa [15] and the auxiliary particle filter of Pitt and Shephard [20]. It should be kept in mind, however, that the successful formulation of any algorithm as an instance of αSMC does not necessarily imply that the assumptions (B), (B^+) or (B^{++}) hold, and the validity of Theorems 1 and 2 is in that sense, of course, not automatic.

Appendix

Lemma 4. *If for every $n \geq 0$, α_n is chosen according to the ESS thresholding rule* (13)*, then* (A2) *is satisfied.*

Proof. The proof is by induction. To initialize, we have at rank $n = 0$,

$$\alpha_0 := \begin{cases} \mathbf{1}_{1/N}, & \text{if } \dfrac{(N^{-1}\sum_i W_0^i g_0(\zeta_0^i))^2}{N^{-1}\sum_i (W_0^i g_0(\zeta_0^i))^2} < \tau, \\ \mathrm{Id}, & \text{otherwise,} \end{cases} \tag{53}$$

noting that by definition $W_0^i = 1$, we find that the entries of α_0 are measurable w.r.t. $\mathcal{F}_0$. For the induction hypothesis, suppose that for some $n \geq 0$ and all $p \leq n$, the entries of α_p are measurable w.r.t. $\mathcal{F}_n$. It follows immediately from Lemma 1, equation (10), that $\{W_{n+1}^i\}_{i\in[N]}$ are all measurable w.r.t. $\mathcal{F}_{n+1}$, and it follows from (13) applied at rank $n+1$ that the entries of α_{n+1} are measurable w.r.t. $\mathcal{F}_{n+1}$, and hence the induction hypothesis holds at rank $n+1$. This completes the proof. □

Resampling times description of the ARPF. In order to derive expressions for π_n^N and Z_n^N in the case of the ARPF, define a family of random sets $\{\sigma_n; n \geq 1\}$, and random times $\{T_n; n \geq 1\}$ as follows

$$\begin{aligned} \sigma_n &:= \{m; 1 \leq m \leq n \text{ and } \alpha_{m-1} = \mathbf{1}_{1/N}\}, \\ T_n &:= \max(\sigma_n), \end{aligned} \tag{54}$$

with $T_n := 0$ on the event $\{\sigma_n = \varnothing\}$. Intuitively, T_n can be thought of as the last resampling time before n. Then by construction, using the recursive definition of W_n^i in (5), and (54), we have on the event $\{\sigma_n \neq \varnothing\}$,

$$\begin{aligned} W_{T_n}^i &= \sum_j \alpha_{T_n-1}^{ij} W_{T_n-1}^j g_{T_n-1}\big(\zeta_{T_n-1}^j\big) \\ &= \frac{1}{N}\sum_j W_{T_n-1}^j g_{T_n-1}\big(\zeta_{T_n-1}^j\big) =: \widetilde{W}_n, \qquad n \geq 1, \end{aligned} \tag{55}$$

which is independent of i. On the event $\{\sigma_n = \varnothing\}$, define $\widetilde{W}_n := 1$.

On the event $\{\sigma_n \neq \varnothing\} \cap \{T_n = n\}$, we trivially have $W_n^i = W_{T_n}^i = \widetilde{W}_n$, by (55). On the event $\{\sigma_n \neq \varnothing\} \cap \{T_n < n\}$, applying equation (9) of Lemma 1 with $p = T_n$, and (55), yields

$$\begin{aligned} W_n^{i_n} &= \sum_{(i_{T_n},\ldots,i_{n-1})\in[N]^{n-T_n}} W_{T_n}^{i_{T_n}} \prod_{q=T_n}^{n-1} g_q\big(\zeta_q^{i_q}\big)\alpha_q^{i_{q+1}i_q} \\ &= \widetilde{W}_n \sum_{(i_{T_n},\ldots,i_{n-1})\in[N]^{n-T_n}} \prod_{q=T_n}^{n-1} g_q\big(\zeta_q^{i_q}\big)\mathbb{I}[i_{q+1} = i_q] = \widetilde{W}_n \prod_{p=T_n}^{n-1} g_p\big(\zeta_p^{i_n}\big). \end{aligned}$$

Collecting the above definitions and substituting into (6) gives

$$\pi_n^N = \frac{\sum_i \delta_{\zeta_n^i} \prod_{p=T_n}^{n-1} g_p(\zeta_p^i)}{\sum_i \prod_{p=T_n}^{n-1} g_p(\zeta_p^i)}, \qquad Z_n^N = \widetilde{W}_n \cdot \frac{1}{N}\sum_i \prod_{p=T_n}^{n-1} g_p\big(\zeta_p^i\big),$$

with the convention $\prod_{p=n}^{n-1} g_p(\zeta_p^i) = 1$. Similar elementary calculations can be used to derive expressions for the sampling steps of the ARPF, in the interests of brevity we leave it to the reader to write out the details.

Proofs and auxiliary results for Section 3

The proof of the main result of Section 3, Theorem 1, hinges on a martingale decomposition of errors associated with Z_n^N and $\pi_n^N(\varphi)$. This is the subject of Proposition 1, which we prove below. Our overall approach is inspired by some of the ideas of [8], Chapters 7 and 9, but the path we take and the details are necessarily different since the analysis of [8] does not apply to αSMC in general. The following well-known lemma has been used extensively in the study of sequential Monte Carlo methods and we shall apply it in the proof of Proposition 1.

Lemma 5 ([8], Lemma 7.3.3). *Let $(\mu_i)_{i\geq 1}$ and $(h_i)_{i\geq 1}$ be, respectively, a sequence of probability measures and a sequence of $\mathbb{R}$-valued, measurable functions with finite oscillations on a given measurable space $(E,\mathcal{E})$. Assume that $\mu_i(h_i)=0$ for all $i\geq 1$ and let $(X_i)_{i\geq 1}$ be a sequence of independent random variables with respective distributions $(\mu_i)_{i\geq 1}$. Then for any $r\geq 1$,*

$$\sqrt{N}\mathbb{E}\left[\left|N^{-1}\sum_{i=1}^{N}h_i(X_i)\right|^r\right]^{1/r}\leq d(r)^{1/r}\sqrt{N^{-1}\sum_{i=1}^{N}\big[\operatorname{osc}(h_i)\big]^2},$$

where for $r\geq 1$,

$$d(2r)=2^{-r}\frac{(2r)!}{r!},\qquad d(2r-1)=\frac{2^{-(r-1/2)}}{(r-1/2)^{1/2}}\frac{(2r-1)!}{(r-1)!}.$$

Proof of Proposition 1. Applying the identities $\beta_{p-1,n}^{i_{p-1}}=\sum_{i_p}\beta_{p,n}^{i_p}\alpha_{p-1}^{i_pi_{p-1}}$, see (22), and $\overline{W}_p^{i_p}=\sum_{i_{p-1}}\alpha_{p-1}^{i_pi_{p-1}}\overline{W}_{p-1}^{i_{p-1}}\overline{Q}_p(1)(\zeta_{p-1}^{i_{p-1}})$, see (5), (32), (34); with the conventions $\alpha_{-1}:=\mathrm{Id}$, and $\overline{\Gamma}_{-1}^N\overline{Q}_{-1,n}(\varphi)=\overline{W}_{-1}^{i}\overline{Q}_{-1,n}(\varphi)(\zeta_{-1}^{i}):=\mu_0\overline{Q}_{0,n}(\varphi)=\pi_n(\varphi)$, we have

$$\begin{aligned}
&\overline{\Gamma}_n^N(\varphi)-\pi_n(\varphi)\\
&\quad=\sum_{p=0}^{n}\big[\overline{\Gamma}_{p,n}^N\overline{Q}_{p,n}(\varphi)-\overline{\Gamma}_{p-1,n}^N\overline{Q}_{p-1,n}(\varphi)\big]\\
&\quad=\sum_{p=0}^{n}\left[\sum_{i_p}\beta_{p,n}^{i_p}\overline{W}_p^{i_p}\overline{Q}_{p,n}(\varphi)\big(\zeta_p^{i_p}\big)-\sum_{i_{p-1}}\sum_{i_p}\beta_{p,n}^{i_p}\alpha_{p-1}^{i_pi_{p-1}}\overline{W}_{p-1}^{i_{p-1}}\overline{Q}_{p-1,n}(\varphi)\big(\zeta_{p-1}^{i_{p-1}}\big)\right] \qquad (56)\\
&\quad=\sum_{p=0}^{n}\sum_{i_p}\beta_{p,n}^{i_p}\overline{W}_p^{i_p}\left[\overline{Q}_{p,n}(\varphi)\big(\zeta_p^{i_p}\big)-\frac{\sum_{i_{p-1}}\alpha_{p-1}^{i_pi_{p-1}}\overline{W}_{p-1}^{i_{p-1}}\overline{Q}_{p-1,n}(\varphi)(\zeta_{p-1}^{i_{p-1}})}{\overline{W}_p^{i_p}}\right]\\
&\quad=\sum_{p=0}^{n}\sum_{i_p}\beta_{p,n}^{i_p}\overline{W}_p^{i_p}\Delta_{p,n}^{i_p}=N^{-1/2}\sum_{k=1}^{(n+1)N}\xi_k^N.
\end{aligned}$$

Each ξ_k^N is measurable w.r.t. $\mathcal{F}^{(k)}$ because, using Corollary 1, (A2) and (B) we have that for any $k = 1, \ldots, (n+1)N$, if we set $p := \lfloor (k-1)/N \rfloor$ and $i := k - pN$, the quantity $\Delta_{p,n}^i$ is measurable w.r.t. $\mathcal{F}^{(k)}$ and $\beta_{p,n}^{i_p} \overline{W}_p^{i_p}$ is measurable w.r.t. $\mathcal{F}_{p-1}$.

To verify (36), again use the fact that for any $i \in [N]$ and $0 \le p \le n$, $\beta_{p,n}^i \overline{W}_p^i$ is measurable w.r.t. $\mathcal{F}_{p-1}$, and note that given $\mathcal{F}_{p-1}$, the particles $\{\zeta_p^i\}_{i=1}^N$ are conditionally independent, and distributed as specified in Algorithm 1. Hence for any $k = 1, \ldots, (n+1)N$ and $p := \lfloor (k-1)/N \rfloor$ and $i := k - pN$, we have $\mathbb{E}[\xi_k^N | \mathcal{F}^{(k-1)}] = \sqrt{N} \beta_{p,n}^i \overline{W}_p^i \mathbb{E}[\Delta_{p,n}^i | \mathcal{F}_{p-1}] = 0$.

For the inequality (37), by Minkowski's inequality and (56),

$$\mathbb{E}\big[\big|\overline{\Gamma}_n^N(\varphi) - \pi_n(\varphi)\big|^r\big]^{1/r} \le \sum_{p=0}^{n} \mathbb{E}\big[\big|\overline{\Gamma}_{p,n}^N \overline{Q}_{p,n}(\varphi) - \overline{\Gamma}_{p-1}^N \overline{Q}_{p-1,n}(\varphi)\big|^r\big]^{1/r}. \tag{57}$$

For each term in (57), using the above stated conditional independence and measurability properties, we may apply Lemma 5 to establish the existence of an independent constant $B(r)$, depending only on r and such that

$$\begin{aligned} &\mathbb{E}\big[\big|\overline{\Gamma}_{p,n}^N \overline{Q}_{p,n}(\varphi) - \overline{\Gamma}_{p-1}^N \overline{Q}_{p-1,n}(\varphi)\big|^r | \mathcal{F}_{p-1}\big] \\ &\quad = \mathbb{E}\left[\left|\sum_i \beta_{p,n}^i \overline{W}_p^i \Delta_{p,n}^i\right|^r \middle| \mathcal{F}_{p-1}\right] \\ &\quad \le B(r) \operatorname{osc}\big(\overline{Q}_{p,n}(\varphi)\big)^r \left(\sum_i \big(\beta_{p,n}^i \overline{W}_p^i\big)^2\right)^{r/2}, \end{aligned} \tag{58}$$

almost surely. The proof is completed upon combining this estimate with (57). □

Proof of Theorem 1. For part (1), note

$$\overline{\Gamma}_n^N(1) - \pi_n(1) = \frac{Z_n^N}{Z_n} - 1,$$

then applying Proposition 1 with $\varphi = 1$ and using (35)–(36) gives

$$\mathbb{E}\big[Z_n^N\big] = Z_n.$$

Moving to the proof of part (2), let us assume for now, only (A1), (A2) and (B), but not necessarily (B^+). Define $c_n := \sup_x g_n(x)/\pi_n(g_n)$. Under (A1), we have

$$\operatorname{osc}\big(\overline{Q}_{p,n}(\varphi)\big) \le 2\|\varphi\| \sup_x \overline{Q}_{p,n}(1)(x) \le 2\|\varphi\| \prod_{q=p}^{n-1} c_q < +\infty$$

and also using Lemma 1, (34) and the fact that each α_p is a Markov transition matrix, we obtain

$$0 < \overline{W}_p^{i_p} \leq \sum_{(i_0,\ldots,i_{p-1})\in[N]^p} \prod_{q=0}^{p-1} c_q \alpha_q^{i_{q+1} i_q} = \prod_{q=0}^{p-1} c_q < +\infty.$$

From (37), we then obtain

$$\begin{aligned} &\mathbb{E}\big[\big|\overline{\Gamma}_n^N(\varphi) - \pi_n(\varphi)\big|^r\big]^{1/r} \\ &\quad \leq 2\|\varphi\| B(r)^{1/r} \left(\prod_{p=0}^{n-1} c_p\right) \sum_{p=0}^{n} \Big|\sum_i \big(\beta_{p,n}^i\big)^2\Big|^{1/2} \\ &\quad \leq 2\|\varphi\| B(r)^{1/r} \left(\prod_{p=0}^{n-1} c_p\right) \sum_{p=0}^{n} \Big|\max_{i\in[N]} \beta_{p,n}^i\Big|^{1/2}, \end{aligned} \tag{59}$$

where the final inequality holds because $\{\beta_{p,n}^i\}_{i\in[N]}$ is a probability vector. Then invoking (B^+), the convergence in (24) follows from (59) applied with $\varphi = 1$. For (25), we apply Minkowski's inequality, the fact $|\Gamma_n^N(\varphi)/\,\Gamma_n^N(1)| \leq \|\varphi\|$ and (59) twice to obtain

$$\begin{aligned} \mathbb{E}\big[\big|\pi_n^N(\varphi) - \pi_n(\varphi)\big|^r\big]^{1/r} &\leq \mathbb{E}\big[\big|\overline{\Gamma}_n^N(\varphi) - \pi_n(\varphi)\big|^r\big]^{1/r} \\ &\quad + \mathbb{E}\left[\left|\frac{\Gamma_n^N(\varphi)}{\Gamma_n^N(1)}\right|^r \big|\overline{\Gamma}_n^N(1) - 1\big|^r\right]^{1/r} \\ &\leq 4\|\varphi\|\big[B(r)\big]^{1/r} \left(\prod_{p=0}^{n-1} c_p\right) \sum_{p=0}^{n} \Big|\max_{i\in[N]} \beta_{p,n}^i\Big|^{1/2}. \end{aligned} \tag{60}$$

The convergence in probability then follows from Markov's inequality, completing the proof of part (2).

For part (3), under (B^{++}) we have $\beta_{p,n}^i = 1/N$, and therefore $|\max_{i\in[N]} \beta_{p,n}^i|^{1/2} = N^{-1/2}$. Substituting this into (59) with $\varphi = 1$, and into (60), gives (26)–(27). The almost sure convergence follows from the Borel–Cantelli lemma. □

Proofs and auxiliary results for Section 4

Proof of Proposition 2. The proof follows a similar line of argument to [9], Proof of Theorem 16.4.1, but applies to a more general algorithm than considered there. To start, we apply Proposition 1, equation (35) with $\varphi = 1$ and (36), we obtain

$$\mathbb{E}\left[\left(\frac{Z_n^N}{Z_n} - 1\right)^2\right] = \sum_{p=0}^{n} \sum_{i_p} \mathbb{E}\big[\big(\beta_{p,n}^{i_p} \overline{W}_p^{i_p} \Delta_{p,n}^{i_p}\big)^2\big].$$

Under (B^{++}) we have $\beta_{p,n}^{i_p} = 1/N$, then using the other hypotheses of the proposition and noting $\operatorname{osc}(Q_{n,n}(1)) = \operatorname{osc}(1) = 0$, we have for $n \geq 1$,

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{Z_n^N}{Z_n} - 1\right)^2\right] &= \sum_{p=0}^{n}\sum_i \mathbb{E}\left[\frac{1}{N^2}\left(\overline{W}_p^i\right)^2\left(\Delta_{p,n}^i\right)^2\right] \\
&\leq \sum_{p=0}^{n-1} \operatorname{osc}\left(\overline{Q}_{p,n}(1)\right)^2 \mathbb{E}\left[\frac{1}{N^2}\sum_i \left(\overline{W}_p^i\right)^2\right] \\
&= \sum_{p=0}^{n-1} \operatorname{osc}\left(\overline{Q}_{p,n}(1)\right)^2 \frac{1}{N}\mathbb{E}\left[\frac{1}{\mathcal{E}_p^N}\left(\frac{1}{N}\sum_i \overline{W}_p^i\right)^2\right] \\
&\leq \sum_{p=0}^{n-1} \frac{\operatorname{osc}(\overline{Q}_{p,n}(1))^2}{N\tau_p}\mathbb{E}\left[\left(\frac{1}{N}\sum_i \overline{W}_p^i - 1\right)^2 + 1\right] \\
&= \sum_{p=0}^{n-1} \frac{\operatorname{osc}(\overline{Q}_{p,n}(1))^2}{N\tau_p}\left(\mathbb{E}\left[\left(\frac{Z_p^N}{Z_p} - 1\right)^2\right] + 1\right),
\end{aligned}$$

where last two lines use $N^{-1}\sum_i \overline{W}_p^i = Z_p^N/Z_p$ and by Theorem 1, $\mathbb{E}[Z_p^N] = Z_p$. □

Proof of Proposition 3. First, note that by the same arguments as in the proof of Proposition 1, equation (58), we have for any $\phi \in \mathcal{L}$, $0 \leq p \leq n$,

$$\begin{aligned}
&\mathbb{E}\left[\left|\Gamma_{p,n}^N Q_{p,n}(\phi) - \Gamma_{p-1}^N Q_{p-1,n}(\phi)\right|^r | \mathcal{F}_{p-1}\right] \\
&\quad \leq B(r)\operatorname{osc}\left(Q_{p,n}(\phi)\right)^r \left(\sum_i \left(\beta_{p,n}^i W_p^i\right)^2\right)^{r/2},
\end{aligned} \tag{61}$$

with the convention $\Gamma_{-1}^N Q_{-1,n}(\phi) = \gamma_n(\phi)$.

For the remainder of the proof, fix $\varphi \in \mathcal{L}$ arbitrarily, and set $\bar{\varphi} := \varphi - \pi_n(\varphi)$. Defining

$$D_{p,n}^N := \frac{\Gamma_{p,n}^N Q_{p,n}(\bar{\varphi})}{\Gamma_{p,n}^N Q_{p,n}(1)}, \qquad 0 \leq p \leq n,$$

and then noting

$$D_{n,n}^N = \frac{\Gamma_n^N(\bar{\varphi})}{\Gamma_n^N(1)} = \pi_n^N(\varphi) - \pi_n(\varphi),$$

we shall focus on the decomposition:

$$\pi_n^N(\varphi) - \pi_n(\varphi) = D_{0,n}^N + \sum_{p=1}^{n} D_{p,n}^N - D_{p-1,n}^N, \tag{62}$$

with the convention that the summation is zero when $n = 0$.

For $1 \leq p \leq n$, write

$$D_{p,n}^N - D_{p-1,n}^N = T_{p,n}^{(N,1)} + T_{p,n}^{(N,2)},$$

where

$$T_{p,n}^{(N,1)} := \frac{1}{\Gamma_{p,n}^N Q_{p,n}(1)} \left[\Gamma_{p,n}^N Q_{p,n}(\bar{\varphi}) - \Gamma_{p-1,n}^N Q_{p-1,n}(\bar{\varphi}) \right],$$

$$T_{p,n}^{(N,2)} := \frac{\Gamma_{p-1,n}^N Q_{p-1,n}(\bar{\varphi})}{\Gamma_{p-1,n}^N Q_{p-1,n}(1)} \frac{[\Gamma_{p-1,n}^N Q_{p-1,n}(1) - \Gamma_{p,n}^N Q_{p,n}(1)]}{\Gamma_{p,n}^N Q_{p,n}(1)}.$$

We have the estimates

$$\frac{\operatorname{osc}(Q_{p,n}(\phi))}{\inf_x Q_{p,n}(1)(x)} \leq 2\delta_{p,n} \left\| P_{p,n}(\phi) \right\| \tag{63}$$

(which is finite under assumption (C) – see also (39)), and

$$\left| \frac{\Gamma_{p-1,n}^N Q_{p-1,n}(\phi)}{\Gamma_{p-1,n}^N Q_{p-1,n}(1)} \right| \leq \left\| P_{p-1,n}(\phi) \right\|. \tag{64}$$

Applying (61) with $\phi = \bar{\varphi}$, using (63) and noting that $\Gamma_{p,n}^N(1)$ is measurable w.r.t. $\mathcal{F}_{p-1}$, we obtain

$$\mathbb{E}\left[\left| T_{p,n}^{(N,1)} \right|^r | \mathcal{F}_{p-1} \right]^{1/r} \leq B(r)^{1/r} 2\delta_{p,n} \frac{\| P_{p,n}(\bar{\varphi}) \|}{\Gamma_{p,n}^N(1)} \left(\sum_i \left(\beta_{p,n}^i W_p^i \right)^2 \right)^{1/2}.$$

Applying (61) with $\phi = 1$, using (64) and the same measurability condition, we obtain

$$\mathbb{E}\left[\left| T_{p,n}^{(N,2)} \right|^r | \mathcal{F}_{p-1} \right]^{1/r} \leq B(r)^{1/r} 2\delta_{p,n} \frac{\| P_{p-1,n}(\bar{\varphi}) \|}{\Gamma_{p,n}^N(1)} \left(\sum_i \left(\beta_{p,n}^i W_p^i \right)^2 \right)^{1/2}.$$

Therefore, via Minkowski's inequality and using

$$\left\| P_{p-1,n}(\bar{\varphi}) \right\| = \left\| Q_{p-1,n}(\bar{\varphi}) / Q_{p-1,n}(1) \right\| = \left\| Q_p Q_{p,n}(\bar{\varphi}) / Q_{p-1,n}(1) \right\| \leq \left\| P_{p,n}(\bar{\varphi}) \right\|,$$

we have

$$\mathbb{E}\left[\left| D_{p,n}^N - D_{p-1,n}^N \right|^r \right]^{1/r} \leq B(r)^{1/r} 4\delta_{p,n} \left\| P_{p,n}(\bar{\varphi}) \right\| \mathbb{E}\left[\left| \mathcal{C}_{p,n}^N \right|^r \right]^{1/r}. \tag{65}$$

For the remaining term, $D_{0,n}^N$, we have

$$D_{0,n}^N = \frac{1}{\Gamma_{0,n}^N Q_{0,n}(1)} \left(\Gamma_{0,n}^N Q_{0,n}(\bar{\varphi}) - \gamma_n(\bar{\varphi}) \right),$$

where the final equality holds since $\gamma_n(\bar{\varphi}) = \gamma_n(\varphi) - \gamma_n(1)\pi_n(\varphi) = 0$. Using (61) and (63) in a similar fashion to above, we obtain

$$\mathbb{E}\big[\big|D_{0,n}^N\big|^r\big]^{1/r} \leq B(r)^{1/r} 2\delta_{0,n} \big\| P_{0,n}(\bar{\varphi}) \big\| \mathbb{E}\big[\big|\mathcal{C}_{0,n}^N\big|^r\big]^{1/r}. \tag{66}$$

The proof is complete upon using Minkowski's inequality to bound the moments of (62) using (65) and (66). □

Lemma 6 ([5], Corollary 5.2). *Suppose that assumptions* (A2) *and* (B) *hold. If*

$$\sup_{n\geq 1} \mathbb{E}\left[\left(\frac{Z_n^N}{Z_n}\right)^2\right]^{1/n} \leq 1 + \frac{c_1}{N\tau}, \tag{67}$$

then

$$N\tau \geq nc_1 \quad \Longrightarrow \quad \mathbb{E}\left[\left(\frac{Z_n^N}{Z_n} - 1\right)^2\right] \leq \frac{2c_1 n}{N\tau}.$$

Proof. Under (A2) and (B), we have by Theorem 1 that $\mathbb{E}[Z_n^N] = Z_n$. The hypothesis (67) can then be stated equivalently as

$$\mathbb{E}\left[\left(\frac{Z_n^N}{Z_n} - 1\right)^2\right] \leq \left(1 + \frac{c_1}{N\tau}\right)^n - 1 \qquad \forall n \geq 1.$$

Using the fact that $\log(1+x) \leq x$ for any $x \geq 0$ and $\mathrm{e}^x \leq 1 + 2x$ for any $x \in [0, 1]$, we conclude that

$$\left(1 + \frac{c_1 n}{N\tau}\right)^n - 1 = \exp\left[n \log\left(1 + \frac{c_1}{N\tau}\right)\right] - 1 \leq \exp\left(\frac{c_1 n}{N\tau}\right) - 1 \leq \frac{2c_1 n}{N\tau}$$

for any $N\tau \geq c_1 n$. □

Proofs for Section 5

Proof of Lemma 3. Label the vertices of the graph corresponding to A arbitrarily with the integers $[N]$. Let $s \geq 1$ be the number of connected components of this graph. Then for each $\ell \in [s]$ let $B(\ell)$ be the set of labels of the ℓth connected component. Since A is a B-matrix, each connected component is complete, so we have for any $\ell \in [s]$ and $i \in B(\ell)$,

$$W_n^i = \sum_j \alpha_{n-1}^{ij} W_{n-1}^j g_{n-1}\big(\zeta_{n-1}^j\big) = \big|B(\ell)\big|^{-1} \sum_{j \in B(\ell)} W_{n-1}^j g_{n-1}\big(\zeta_{n-1}^j\big). \tag{68}$$

The complexity of calculating W_n^i is thus $\mathrm{O}(|B(\ell)|)$, and since $W_n^i = W_n^j$ for all $i, j \in B(\ell)$, the complexity of calculating $\{W_n^i\}_{i\in[N]}$ is $\mathrm{O}(\sum_{\ell\in[s]} |B(\ell)|) = \mathrm{O}(N)$. Arguing similarly to (68),

with $\alpha_{n-1} = A$ we find that under Algorithm 1, for each $\ell \in [s]$, the $\{\zeta_n^i\}_{i\in B(\ell)}$ are conditionally i.i.d. according to

$$\frac{\sum_{j\in B(\ell)} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j) f(\zeta_{n-1}^j, \cdot)}{\sum_{j\in B(\ell)} W_{n-1}^j g_{n-1}(\zeta_{n-1}^j)}. \tag{69}$$

By the same arguments used in [4] to address the BPF, drawing $|B(\ell)|$ samples from (69) can be achieved at $O(|B(\ell)|)$ complexity, and thus the overall complexity of the sampling part of Algorithm 1 is $O(\sum_{\ell\in[s]} |B(\ell)|) = O(N)$. □

Proof of Proposition 4. We prove (51) by induction. We have

$$\mathbb{W}_0^i = W_{n-1}^i g_{n-1}\big(\zeta_{n-1}^i\big) = \sum_{j\in B(0,i)} W_{n-1}^j g_{n-1}\big(\zeta_{n-1}^j\big)$$

and, when (51) holds at rank k, we have at rank $k+1$,

$$\begin{aligned}
\mathbb{W}_{k+1}^i &= \mathbb{W}_k^{\mathcal{I}_k(2i-1)}/2 + \mathbb{W}_k^{\mathcal{I}_k(2i)}/2 \\
&= 2^{-(k+1)} \sum_{j\in B(k,\mathcal{I}_k(2i-1))} W_{n-1}^j g_{n-1}\big(\zeta_{n-1}^j\big) \\
&\quad + 2^{-(k+1)} \sum_{j\in B(k,\mathcal{I}_k(2i))} W_{n-1}^j g_{n-1}\big(\zeta_{n-1}^j\big) \\
&= 2^{-(k+1)} \sum_{j\in B(k,\mathcal{I}_k(2i-1))\cup B(k,\mathcal{I}_k(2i))} W_{n-1}^j g_{n-1}\big(\zeta_{n-1}^j\big) \\
&= 2^{-(k+1)} \sum_{j\in B(k+1,i)} W_{n-1}^j g_{n-1}\big(\zeta_{n-1}^j\big).
\end{aligned}$$

Finally, for any $i \in [N/2^k]$ and $j \in B(k,i)$

$$W_n^j = \sum_\ell \alpha_{n-1}^{i\ell} W_{n-1}^\ell g_{n-1}\big(\zeta_{n-1}^\ell\big) = 2^{-k} \sum_{\ell\in B(k,i)} W_{n-1}^\ell g_{n-1}\big(\zeta_{n-1}^\ell\big) = \mathbb{W}_k^i,$$

which establishes (51)–(52).

No matter what adaptation rule of Table 1 is used, each quantity $\{B(k,i)\}_{i\in[N/2^k]}$ obtained by Algorithm 4 is, by construction, a partition of $[N]$ and thus the α_{n-1} output by Algorithm 4 is a B-matrix. Noting that a B-matrix always admits the uniform distribution on $[N]$ as an invariant distribution, we have for any B-matrix, say A, the identity $\overline{\mathbb{W}}_0 = N^{-1}\sum_i \sum_j A^{ij} W_{n-1}^i g_{n-1}(\zeta_{n-1}^i)$ and so upon termination of the "while" loop in Algorithm 4, $\mathcal{E} = \mathcal{E}_n^N$ and hence $\mathcal{E}_n^N \geq \tau$ always.

For the Simple and Random adaptation rules, the worst case complexity of Algorithm 4 is as follows. The part of the algorithm preceding the "while" loop is $O(N)$. The complexity of iteration k of the "while" loop is $O(N/2^k)$, the worst case is when the loop terminates with

$k = m$, in which case the complexity of the "while" loop is $\mathrm{O}(\sum_{k=0}^{m} N/2^k)$, thus the overall complexity is no more than $\mathrm{O}(N)$.

For the Greedy procedure, the sort operation required to obtain $\mathcal{I}_k$ is of complexity $\mathrm{O}(N/2^k \log_2(N/2^k))$, and so in the worst case, the complexity of the "while" loop is of the order

$$t(N) := \sum_{k=0}^{m} \frac{N}{2^k} \log_2\left(\frac{N}{2^k}\right),$$

or expressed recursively, $t(N) = t(N/2) + N \log_2 N$, and $t(2) = 2$. A simple induction shows that this recursion has solution $t(N) = 2[1 + N(\log_2 N - 1)]$, hence the overall worst case complexity of the "while" loop is $\mathrm{O}(N \log_2 N)$. The proof is complete since by Lemma 3, the complexity of operations in Algorithm 1 other than line $(\star)$ is $\mathrm{O}(N)$. □

Acknowledgements

The authors would like to thank the Associate Editor and a referee for helpful comments and corrections. The first and third authors were supported by EPSRC grant EP/K023330/1.

References

[1] Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 269–342. MR2758115

[2] Arnaud, E. and Le Gland, F. (2009). SMC with adaptive resampling: Large sample asymptotics. In *IEEE/SP 15th Workshop on Statistical Signal Processing* 481–484.

[3] Bolić, M., Djurić, P.M. and Hong, S. (2005). Resampling algorithms and architectures for distributed particle filters. *IEEE Trans. Signal Process.* **53** 2442–2450. MR2190405

[4] Carpenter, J., Clifford, P. and Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proceedings – Radar, Sonar and Navigation* **146** 2–7.

[5] Cérou, F., Del Moral, P. and Guyader, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 629–649. MR2841068

[6] Chen, T., Schön, T.B., Ohlsson, H. and Ljung, L. (2011). Decentralized particle filter with arbitrary state decomposition. *IEEE Trans. Signal Process.* **59** 465–478. MR2790606

[7] Crisan, D. and Obanubi, O. (2012). Particle filters with random resampling times. *Stochastic Process. Appl.* **122** 1332–1368. MR2914755

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

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

[10] Del Moral, P., Doucet, A. and Jasra, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 411–436. MR2278333

[11] Del Moral, P., Doucet, A. and Jasra, A. (2012). On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli* **18** 252–278. MR2888706

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

[13] Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *Radar and Signal Processing*, *IEE Proceedings F* **140** 107–113. IET.
[14] Johansen, A.M., Whiteley, N. and Doucet, A. (2012). Exact approximation of Rao–Blackwellised particle filters. In 16*th IFAC Symposium on System Identification* **16** 488–493.
[15] Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5** 1–25. MR1380850
[16] Künsch, H.R. (2013). Particle filters. *Bernoulli* **19** 1391–1403. MR3102556
[17] Lee, A., Yau, C., Giles, M.B., Doucet, A. and Holmes, C.C. (2010). On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods. *J. Comput. Graph. Statist.* **19** 769–789.
[18] Le Gland, F. and Oudjane, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Probab.* **14** 144–187. MR2023019
[19] Liu, J.S. and Chen, R. (1995). Blind deconvolution via sequential imputations. *J. Amer. Statist. Assoc.* **90** 567–576.
[20] Pitt, M.K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. MR1702328
[21] Vergé, C., Dubarry, C., Del Moral, P. and Moulines, E. (2013). On parallel implementation of sequential Monte Carlo methods: The island particle model. *Stat. Comput.* **23** 1–18.
[22] Whiteley, N. (2013). Stability properties of some particle filters. *Ann. Appl. Probab.* **23** 2500–2537. MR3127943
[23] Whiteley, N. and Lee, A. (2014). Twisted particle filters. *Ann. Statist.* **42** 115–141. MR3178458

Received September 2013 and revised July 2014