Methodol Comput Appl Probab (2017) 19:799–818
DOI 10.1007/s11009-016-9513-8

An Algorithm for Approximating the Second Moment of the Normalizing Constant Estimate from a Particle Filter

Svetoslav Kostov[1] · Nick Whiteley[1]

Received: 6 February 2016 / Revised: 17 August 2016 /
Accepted: 24 August 2016 / Published online: 27 September 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract We propose a new algorithm for approximating the non-asymptotic second moment of the marginal likelihood estimate, or normalizing constant, provided by a particle filter. The computational cost of the new method is $O(M)$ per time step, independently of the number of particles N in the particle filter, where M is a parameter controlling the quality of the approximation. This is in contrast to $O(MN)$ for a simple averaging technique using M i.i.d. replicates of a particle filter with N particles. We establish that the approximation delivered by the new algorithm is unbiased, strongly consistent and, under standard regularity conditions, increasing M linearly with time is sufficient to prevent growth of the relative variance of the approximation, whereas for the simple averaging technique it can be necessary to increase M exponentially with time in order to achieve the same effect. This makes the new algorithm useful as part of strategies for estimating Monte Carlo variance. Numerical examples illustrate performance in the context of a stochastic Lotka–Volterra system and a simple AR(1) model.

Keywords Marginal likelihood · Normalizing constant · Hidden Markov model · Particle filter

Mathematics Subject Classification (2010) 93E11 · 65C05

✉ Svetoslav Kostov
sk12540@bristol.ac.uk

Nick Whiteley
nick.whiteley@bristol.ac.uk

[1] School of Mathematics, University Walk, Bristol, BS8 1TW, UK

1 Introduction

Particle filters, also known as Sequential Monte Carlo (SMC) methods (Doucet et al. 2001), are used across a variety of disciplines including systems biology, econometrics, neuroscience and signal processing, to perform approximate inferential calculations in general state-space Hidden Markov Models (HMM) and in particular, provide an unbiased estimate of the marginal likelihood. Recent application areas of these techniques include for example, systems biology (Golightly and Wilkinson 2011; Golightly et al. 2015), where the calculation of the marginal likelihood (ML) plays an important role in the estimation of the parameters of stochastic models of biochemical networks. Estimation of the marginal likelihood also features centrally in Particle Markov Chain Monte Carlo methods (Andrieu et al. 2010).

In the present paper we address the problem of approximating the non-asymptotic second moment of the particle filter estimate of the marginal likelihood, henceforth for brevity "the second moment". As part of strategies to estimate Monte Carlo variance, this allows one to report a numerical measure of the reliability of the particle filter estimate. Our contributions are to introduce a new particle "Pairs algorithm" and prove that it unbiasedly and consistently approximates the second moment. We also establish, under regularity conditions, a linear-in-time bound on the relative variance of the approximation to the second moment, and illustrate through a simple calculation and numerical simulations, that the Pairs algorithm performs more reliably than a default strategy which uses independent copies of the particle filter. In order to discuss the connections between our work and the existing literature, we first need to introduce some notation and definitions.

A HMM is a process $(X_n, Y_n)_{n\geqslant 0}$, where $(X_n)_{n\geqslant 0}$, called the signal process, is a Markov chain with state space X, initial distribution π_0 and transition kernel f. Each of the observations $Y_n \in \mathsf{Y}$, is conditionally independent of the rest of the signal process given X_n, with conditional distribution, $g(X_n, \cdot)$, where g is a probability kernel from X to Y. The HMM can be represented as:

$$\begin{aligned} X_0 \sim \pi_0(\cdot), \quad X_n \mid X_{n-1} &\sim f(X_{n-1}, \cdot), \quad n \geqslant 1 \\ Y_n \mid X_n &\sim g(X_n, \cdot), \quad n \geqslant 0. \end{aligned} \tag{1}$$

We consider a fixed observation sequence $(y_n)_{n\geqslant 0}$, assume that g admits a density $g(x, y)$ w.r.t. to some dominating measure and write for brevity $g_n(x) = g(x, y_n)$. For simplicity we also assume throughout that for all $n \geqslant 0$, $\sup_x g_n(x) < +\infty$ and $g_n(x) > 0$, $\forall x \in \mathsf{X}$. We then define the sequence of distributions $(\pi_n)_{n\geqslant 1}$, called prediction filters, as

$$\pi_{n+1}(A) := \frac{\int_{\mathsf{X}} \pi_n(dx) g_n(x) f(x, A)}{\int_{\mathsf{X}} \pi_n(dx) g_n(x)}, \quad \forall A \in \mathcal{X},\ n \geqslant 0,$$

where $\mathcal{X}$ is the σ-algebra associated with the space X, and the sequence

$$(Z_n)_{n\geqslant 0}, \quad Z_0 := \int_{\mathsf{X}} g_0(x)\pi_0(dx), \quad Z_n := Z_{n-1} \int_{\mathsf{X}} g_n(x)\pi_n(dx), \quad n \geqslant 1. \tag{2}$$

The interpretation of these definitions is the following: π_{n+1} is the distribution of $X_{n+1} \mid Y_{0:n} = y_{0:n}$, where for any sequence $(a_n)_{n\geqslant 0}$ we write $a_{p:q} = (a_p, \ldots, a_q)$, and Z_n is the marginal likelihood of the first $n+1$ observations $y_{0:n}$. In many cases of interest, the distributions π_n and constants Z_n cannot be computed exactly, and numerical approximations are needed. A particle filter, shown in Algorithm 1, provides such approximations, denoted respectively π_n^N and Z_n^N. In Algorithm 1 q_0 and q_n, $n \geqslant 1$ are respectively a distribution and Markov kernels on X, which may depend on the observations sequence $(y_n)_{n\geqslant 0}$,

but this dependence is suppressed from the notation. We assume throughout the rest of the paper that $\pi_0(\cdot)$, $f(x,\cdot)$ and $q_0(\cdot)$ and $q_n(x,\cdot)$ admit a density w.r.t. to some common dominating measure dx, and with a slight abuse of notation, the corresponding densities are denoted by $\pi_0(x)$, $f(x,x')$, $q_0(x)$ and $q_n(x,x')$.

Algorithm 1 SMC algorithm for estimating Z_n using N particles

Initialization

- Sample $\{X_0^i\}_{i=1}^N \overset{i.i.d.}{\sim} q_0(\cdot)$
- Compute weights $\{W_0^i\}_{i=1}^N$ according to $W_0^i = \dfrac{g_0(X_0^i)\pi_0(X_0^i)}{q_0(X_0^i)}$

  normalize, $\widetilde{W}_0^i = \dfrac{W_0^i}{\sum_{k=1}^N W_0^k}$, and set $Z_0^N = \dfrac{1}{N}\sum_{i=1}^N W_0^i$
- Resample conditionally i.i.d. draws from $\{X_0^i\}_{i=1}^N$ using the normalized weights $\{\widetilde{W}_0^i\}_{i=1}^N$ to obtain a set of equally-weighted particles $\{\underline{X}_0^i\}_{i=1}^N$

For $n \geqslant 1$:

- For each i, set $X_{n-1}^i = \underline{X}_{n-1}^i$
- For each i, sample $X_n^i \sim q_n(X_{n-1}^i,\cdot)$, compute weights $W_n^i = \dfrac{g_n(X_n^i)f(X_{n-1}^i,X_n^i)}{q_n(X_{n-1}^i,X_n^i)}$,

  normalize, $\widetilde{W}_n^i = \dfrac{W_n^i}{\sum_{k=1}^N W_n^k}$, and set $Z_n^N = Z_{n-1}^N \cdot \left(\dfrac{1}{N}\sum_{i=1}^N W_n^i\right)$
- Resample conditionally i.i.d. draws from $\{X_n^i\}_{i=1}^N$ using the normalized weights $\{\widetilde{W}_n^i\}_{i=1}^N$ to obtain a set of equally-weighted particles $\{\underline{X}_n^i\}_{i=1}^N$

It is well known that Algorithm 1 provides an unbiased estimate of Z_n, i.e. $\mathbb{E}\big[Z_n^N\big] = Z_n$. A detailed account of this fact is given in (Del Moral 2004, Ch. 9). The main contribution of the present paper is to propose and study a new method to approximate $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$. The approximation is delivered by Algorithm 2 – the Pairs algorithm – which we introduce in the next section, and which must be run *in addition* to the particle filter used to estimate Z_n^N. Our main motivation for approximating $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ is to calculate $\mathrm{Var}\big[Z_n^N\big]$. In a recent arXiv manuscript (Lee and Whiteley 2015), A. Lee and the second author of the present paper have introduced a method which allows one to unbiasedly approximate $\mathrm{Var}\big[Z_n^N\big]$ using the same single run of the particle filter which delivers Z_n^N. As $N \to \infty$, the method of Lee and Whiteley (2015) allows one to consistently approximate asymptotic variance $\lim_{N\to\infty} N\mathrm{Var}\big[Z_n^N\big]$.

We stress that the Pairs algorithm performs the different task of approximating, for any *fixed* $N \geqslant 2$, the non-asymptotic quantity $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ to arbitrary accuracy controlled by an auxiliary parameter M (this statement is made precise in Theorem 2.1 below). Thus the Pairs algorithm allows one to reliably approximate $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ without requiring that N is large. We shall later illustrate how this property makes the Pairs algorithm useful within strategies for estimating $\mathrm{Var}\big[Z_n^N\big]$.

Moreover in Theorem 2.1 we prove an important result regarding the time dependence of the error of the approximation of $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ delivered by the Pairs algorithm, showing that under standard regularity conditions, it is sufficient to increase M linearly with n to control the relative variance of this approximation. This is in contrast to Lee and Whiteley (2015), who do not provide any results concerning the time-dependence of the errors associated with their estimators.

We note that Chan and Lai (2013) investigated numerical techniques for assessing the asymptotic variance associated with particle estimates of expectations with respect to filtering distributions, but they didn't explore methods for approximating $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$. We also note that Bhadra and Ionides (2014) proposed to approximate $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ using a "meta-model", for purposes of optimizing parameters of the particle filter. Their method amounts to fitting an AR(1) process to the output of the particle filter; it seems difficult to assess the bias of their approach and no proof of consistency is given.

2 Pairs Algorithm

2.1 Outline of How the Algorithm is Derived

The full details of the derivation of the Pairs algorithm are given in Appendix. We now give an account of some of the main ideas behind this derivation. For this some more notation is needed. Let us introduce the nonnegative integral kernels: for $x \in \mathsf{X}$, $y = (y_1, y_2) \in \mathsf{X}^2$,

$$Q_1(x, dy) = \frac{g_0(x)\pi_0(x)}{q_0(x)} q_1(y_1, y_2)\delta_x(dy_1)dy_2, \tag{3}$$

and for $n \geqslant 2$ and $x = (x_1, x_2) \in \mathsf{X}^2$, $y \in \mathsf{X}^2$,

$$Q_n(x, dy) = \frac{g_{n-1}(x_2) f(x_1, x_2)}{q_{n-1}(x_1, x_2)} q_n(y_1, y_2)\delta_{x_2}(dy_1)dy_2. \tag{4}$$

In terms of compositions of these kernels, the lack-of-bias property of the particle filter reads as:

$$\mathbb{E}\left[Z_n^N\right] = \pi_0 Q_1 \cdots Q_n(1). \tag{5}$$

The kernels also encapsulate the main ingredients of the particle filter itself, indeed one may take the point of view that Algorithm 1 is actually derived from the Q_n, in the sense that resampling is performed according to weights given by evaluating the functions

$$Q_1(x, \mathsf{X}^2) = \frac{g_0(x)\pi_0(x)}{q_0(x)}, \quad Q_n(x, \mathsf{X}^2) = \frac{g_{n-1}(x_2) f(x_1, x_2)}{q_{n-1}(x_1, x_2)}, \quad n \geqslant 2, \tag{6}$$

and sampling is performed using the the Markov kernels:

$$\frac{Q_n(x, \cdot)}{Q_n(x, \mathsf{X}^2)}. \tag{7}$$

Now introduce the so–called coalescence operator C which acts on functions $F : \mathsf{X}^2 \times \mathsf{X}^2 \to \mathbb{R}$ as $C(F)(x, y) = F(x, x)$. Cérou et al. (2011) derived the following representation of the second moment of Z_n^N,

$$\mathbb{E}\left[\left(Z_n^N\right)^2\right] = \mathbb{E}\left[\pi_0^{\otimes 2} C_{\epsilon_0} Q_1^{\otimes 2} C_{\epsilon_1} \cdots C_{\epsilon_n} Q_{n+1}^{\otimes 2}(1)\right], \tag{8}$$

where $C_1 := C$, $C_0 := Id$, $\{\epsilon_n\}_{n\geqslant 0}$ is a sequence of i.i.d., $\{0, 1\}$-valued random variables with distribution

$$\mathbb{P}(\epsilon_n = 1) = 1 - \mathbb{P}(\epsilon_n = 0) = \frac{1}{N},$$

and $Q_n^{\otimes 2}$ is the two-fold tensor product of Q_n.

The main idea behind the Pairs algorithm is to identify, using Eq. 8, certain nonnegative kernels $\mathbf{Q}_n^{(N)}$ such that the second moment can be written

$$\mathbb{E}\left[\left(Z_n^N\right)^2\right] = \pi_0^{\otimes 2}\mathbf{Q}_1^{(N)}\cdots\mathbf{Q}_{n+1}^{(N)}(1).$$

The details of these kernels $\mathbf{Q}_n^{(N)}$ are given in the Appendix. Observing the similarity with Eq. 5, to obtain the Pairs algorithm we shall derive a particle algorithm from the weighting functions and Markov kernels which are associated with $\mathbf{Q}_n^{(N)}$ in the same way as Eqs. 6–7 are associated with Q_n, the result being the Pairs algorithm. Results for standard particle filters then transfer to the Pairs algorithm directly, which leads to our Theorem 2.1 below.

2.2 The Algorithm and its Properties

In Algorithm 2 both $N \geqslant 2$ and $M \geqslant 1$ are parameters. The computational cost of Algorithm 2 is $O(M)$ per time step, uniformly in N, and the quantity $\Xi_n^{(N,M)}$ which it delivers can be considered an approximation to $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$, in the sense of Theorem 2.1 below.

Theorem 2.1 *If*

$$\sup_x \frac{g_0(x)\pi_0(x)}{q_0(x)} < +\infty \quad \textit{and} \quad \sup_{x_1,x_2} \frac{g_n(x_2)f(x_1,x_2)}{q_n(x_1,x_2)} < +\infty, \quad \forall n \geqslant 1, \tag{9}$$

then for any $N \geqslant 2$ and $n \geqslant 0$,

$$\mathbb{E}\left[\Xi_n^{(N,M)}\right] = \mathbb{E}\left[\left(Z_n^N\right)^2\right], \quad \forall M \geqslant 1,$$

$$\Xi_n^{(N,M)} \xrightarrow[M\to\infty]{a.s.} \mathbb{E}\left[\left(Z_n^N\right)^2\right].$$

If additionally for each $n \geqslant 0$ there exist constants $0 < w_n^- \leqslant w_n^+ < +\infty$, and for each $n \geqslant 1$, constants $0 < \epsilon_n^- \leqslant \epsilon_n^+ < +\infty$ and a probability measure μ_n such that

$$w_0^- \leqslant g_0(x)\pi_0(x)/q_0(x) \leqslant w_0^+, \quad \forall x, \tag{10}$$

$$w_n^- \leqslant g_n(x_2)f(x_1,x_2)/q_n(x_1,x_2) \leqslant w_n^+, \quad \forall x_1, x_2, n \geqslant 1, \tag{11}$$

$$\epsilon_n^-\mu_n(\cdot) \leqslant q_n(x,\cdot) \leqslant \epsilon_n^+\mu_n(\cdot), \quad \forall x, n \geqslant 1, \tag{12}$$

then for any $N \geqslant 2$ and $n \geqslant 0$,

$$M > \sum_{s=0}^{n+1}\Delta_s \quad \Rightarrow \quad \mathbb{E}\left[\left(\frac{\Xi_n^{(N,M)}}{\mathbb{E}\left[\left(Z_n^N\right)^2\right]} - 1\right)^2\right] \leqslant \frac{4}{M}\sum_{s=0}^{n+1}\Delta_s$$

where $\Delta_s := \left(\frac{w_s^+ w_{s+1}^+ \epsilon_{s+1}^+}{w_s^- w_{s+1}^- \epsilon_{s+1}^-}\right)^2$ is independent of M and N.

Algorithm 2 Pairs algorithm for approximating $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ using M pair particles

Initialization

- Sample pairs $\left\{\check{X}_0^i\right\}_{i=1}^M \overset{i.i.d.}{\sim} q_0(\cdot)$, $\left\{\hat{X}_0^i\right\}_{i=1}^M \overset{i.i.d.}{\sim} q_0(\cdot)$
- Compute weights $\left\{W_0^i\right\}_{i=1}^M$ according to

$$W_0^i = \frac{1}{N}\frac{g_0(\check{X}_0^i)^2\pi_0(\check{X}_0^i)^2}{q_0(\check{X}_0^i)^2} + \left(1-\frac{1}{N}\right)\frac{g_0(\check{X}_0^i)g_0(\hat{X}_0^i)\pi_0(\check{X}_0^i)\pi_0(\hat{X}_0^i)}{q_0(\check{X}_0^i)q_0(\hat{X}_0^i)},$$

  normalize weights $\widetilde{W}_0^i = \dfrac{W_0^i}{\sum_{k=1}^M W_0^k}$ and set $\Xi_0^{(N,M)} = \dfrac{1}{M}\sum_{i=1}^M W_0^i$.
- Resample conditionally i.i.d. draws from $\left\{\check{X}_0^i, \hat{X}_0^i\right\}_{i=1}^M$ using the normalized weights $\left\{\widetilde{W}_0^i\right\}_{i=1}^M$ to obtain a set of equally-weighted particles $\left\{\underline{\check{X}}_0^i, \underline{\hat{X}}_0^i\right\}_{i=1}^M$
- For each i, set $\left(\check{X}_0^i, \hat{X}_0^i\right) = \left(\underline{\check{X}}_0^i, \underline{\hat{X}}_0^i\right)$, compute $p_0^i = \left(1+(N-1)\dfrac{g_0(\hat{X}_0^i)\pi_0(\hat{X}_0^i)q_0(\check{X}_0^i)}{g_0(\check{X}_0^i)\pi_0(\check{X}_0^i)q_0(\hat{X}_0^i)}\right)^{-1}$ and sample $Y_0^i \sim Ber(p_0^i)$. If $Y_0^i = 1$, set $\hat{X}_0^i = \check{X}_0^i$. Sample $\check{X}_1^i \sim q_1(\check{X}_0^i, \cdot)$, $\hat{X}_1^i \sim q_1(\hat{X}_0^i, \cdot)$.

For $n \geqslant 1$:

- Compute weights $\left\{W_n^i\right\}_{i=1}^M$ according to

$$\begin{aligned} W_n^i = {} & \frac{1}{N}\frac{g_n(\check{X}_n^i)^2 f(\check{X}_{n-1}^i, \check{X}_n^i)^2}{q_n(\check{X}_{n-1}^i, \check{X}_n^i)^2} \\ & + \left(1-\frac{1}{N}\right)\frac{g_n(\check{X}_n^i)g_n(\hat{X}_n^i)f(\check{X}_{n-1}^i, \check{X}_n^i)f(\hat{X}_{n-1}^i, \hat{X}_n^i)}{q_n(\check{X}_{n-1}^i, \check{X}_n^i)q_n(\hat{X}_{n-1}^i, \hat{X}_n^i)}, \end{aligned}$$

  normalize, $\widetilde{W}_n^i = \dfrac{W_n^i}{\sum_{k=1}^M W_n^k}$, and set $\Xi_n^{(N,M)} = \Xi_{n-1}^{(N,M)} \cdot \left(\dfrac{1}{M}\sum_{i=1}^M W_n^i\right)$
- Resample conditionally i.i.d. draws from $\left\{\check{X}_{n-1:n}^i, \hat{X}_{n-1:n}^i\right\}_{i=1}^M$ using the normalized weights $\left\{\widetilde{W}_n^i\right\}_{i=1}^M$ to obtain a set of equally-weighted particles $\left\{\underline{\check{X}}_{n-1:n}^i, \underline{\hat{X}}_{n-1:n}^i\right\}_{i=1}^M$
- For each i, set $\left(\check{X}_{n-1:n}^i, \hat{X}_{n-1:n}^i\right) = \left(\underline{\check{X}}_{n-1:n}^i, \underline{\hat{X}}_{n-1:n}^i\right)$, compute $p_n^i = \left(1+(N-1)\dfrac{g_n(\hat{X}_n^i)f(\hat{X}_{n-1}^i, \hat{X}_n^i)q_n(\check{X}_{n-1}^i, \check{X}_n^i)}{g_n(\check{X}_n^i)f(\check{X}_{n-1}^i, \check{X}_n^i)q_n(\hat{X}_{n-1}^i, \hat{X}_n^i)}\right)^{-1}$ and sample $Y_n^i \sim Ber(p_n^i)$. If $Y_n^i = 1$, set $\hat{X}_n^i = \check{X}_n^i$. Sample $\check{X}_{n+1}^i \sim q_{n+1}(\check{X}_n^i, \cdot)$, $\hat{X}_{n+1}^i \sim q_{n+1}(\hat{X}_n^i, \cdot)$.

The proof of Theorem 2.1 is given in Appendix. The conditions in Eqs. 10–12 are fairly standard in the stability theory of particle filters, but are rather strong: they rarely hold when X is an unbounded subset of $\mathbb{R}^d$. Attempting to establish similar results under more realistic conditions, for example via the techniques of Whiteley (2013), seems to be a much more

difficult task, beyond the scope of the present work, and we leave a full investigation of this matter to future research.

2.3 Comparison to Using i.i.d. Replicates of Z_n^N

A natural alternative to $\Xi_n^{(N,M)}$ as an approximation to $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ is to use M i.i.d. replicates $\left\{Z_n^{N,j}\right\}_{j=1}^M$ of Z_n^N and simple averaging,

$$\widetilde{\Xi}_n^{(N,M)} := \frac{1}{M}\sum_{j=1}^{M}\left(Z_n^{N,j}\right)^2. \tag{13}$$

The cost of computing $\widetilde{\Xi}_n^{(N,M)}$ is $O(MN)$ per time step since it involves M copies of Algorithm 1, each using N particles.

To illustrate why $\Xi_n^{(N,M)}$ is to be preferred over $\widetilde{\Xi}_n^{(N,M)}$ in terms of relative variance, consider for simplicity of exposition the case: for $n \geqslant 1$, $q_n(x,\cdot) = f(x,\cdot) = \pi_0(\cdot)$; for $n = 0$, $q_0(\cdot) = \pi_0(\cdot)$; and for $n \geqslant 0$, $g_n(x) = g(x)$. In this case, for all $n \geqslant 0$, we have $\pi_n = \pi_0$ and in Algorithm 1, $\{X_n^i\}_{i=1}^N$ are i.i.d. draws from π_0. Then with $\pi_p^N(g) := N^{-1}\sum_{i=1}^N g(X_p^i)$, $Z_n^N = \prod_{p=0}^n \pi_p^N(g)$, and

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\widetilde{\Xi}_n^{(N,M)}}{\mathbb{E}\left[\left(Z_n^N\right)^2\right]} - 1\right)^2\right] &= \frac{1}{M}\left(\frac{\mathbb{E}\left[\left(Z_n^N\right)^4\right]}{\mathbb{E}\left[\left(Z_n^N\right)^2\right]^2} - 1\right) \\
&= \frac{1}{M}\left(\prod_{p=0}^{n}\frac{\mathbb{E}\left[\pi_p^N(g)^4\right]}{\mathbb{E}\left[\pi_p^N(g)^2\right]^2} - 1\right) \\
&= \frac{1}{M}\left(C^{n+1} - 1\right),
\end{aligned} \tag{14}$$

where $C := \mathbb{E}\left[\pi_0^N(g)^4\right]/\mathbb{E}\left[\pi_0^N(g)^2\right]^2 \geqslant 1$ by Jensen's inequality, with equality holding if and only if $\pi_0^N(g)$ is a.s. constant. So if $\pi_0^N(g)$ exhibits any stochastic variability at all, in the sense that $C > 1$, then M must be scaled exponentially fast with n in order to control (14), cf. the linear-in-n scaling in Theorem 2.1.

3 Numerical Examples

We will illustrate the properties of the Pairs algorithm using two numerical examples. The first, in Section 3.1 is a simple toy example, based on a $AR(1)$ auto-regressive process. The second, in Section 3.2, is a more realistic example involving a Lotka - Volterra system of ODEs, observed in noise. In Section 3.3 we investigated the performance of the pairs algorithm within a strategy for estimating Monte Carlo variance.

Throughout Section 3 we denote by M' a number of pairs used in the pairs algorithm to obtain a reliable, benchmark estimate of the true quantity $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$.

3.1 $AR(1)$ Example

The signal of this model $(X_n)_{n\geqslant 0}$ is an $AR(1)$ process, defined by $X_{n+1} = \alpha X_n + \epsilon_{n+1}$, where we set $\alpha = 0.5$, $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$, $\sigma = 10$. Assume that $g_n(x) = \exp\left(-x^2/100\right)$, $\forall n$. We will also assume that $q_n(x, \cdot) = f(x, \cdot)$, i.e. we will propose using the actual signal density and we will set $q_0 = \pi_0$, given by $X_0 \sim \mathcal{N}(0, \sigma^2/(1-\alpha^2))$, i.e. the process $(X_n)_{n\geqslant 0}$ is stationary a priori.

In Fig. 1 we compare two approaches for estimating $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$: using the Pairs algorithm, and the standard MC approach using i.i.d. replicates as in Eq. 13. We consider two sub–examples: the first one is for comparatively small number of particles $N = 50$, and the second sub–example is with higher number of particles $N = 250$. The plots show $\log(\Xi_n^{(N,M)}) - \log(\Xi_n^{(N,M')})$ for the Pairs algorithm and $\log(\widetilde{\Xi}_n^{(N,\tilde{M})}) - \log(\Xi_n^{(N,M')})$ for the standard MC approach (please refer to Algorithm 2 and Eq. 13). Here we take $M' = 10^6$ so that $\Xi_n^{(N,M')}$ is a reliable, benchmark value of $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$.

In the top left plot of Fig. 1 we have chosen $M = \tilde{M} = 10^4$. For the equal cost plot on the top right we have chosen $M = 10^4$ and $\tilde{M} = 2500$. Here, by "equal cost" we mean

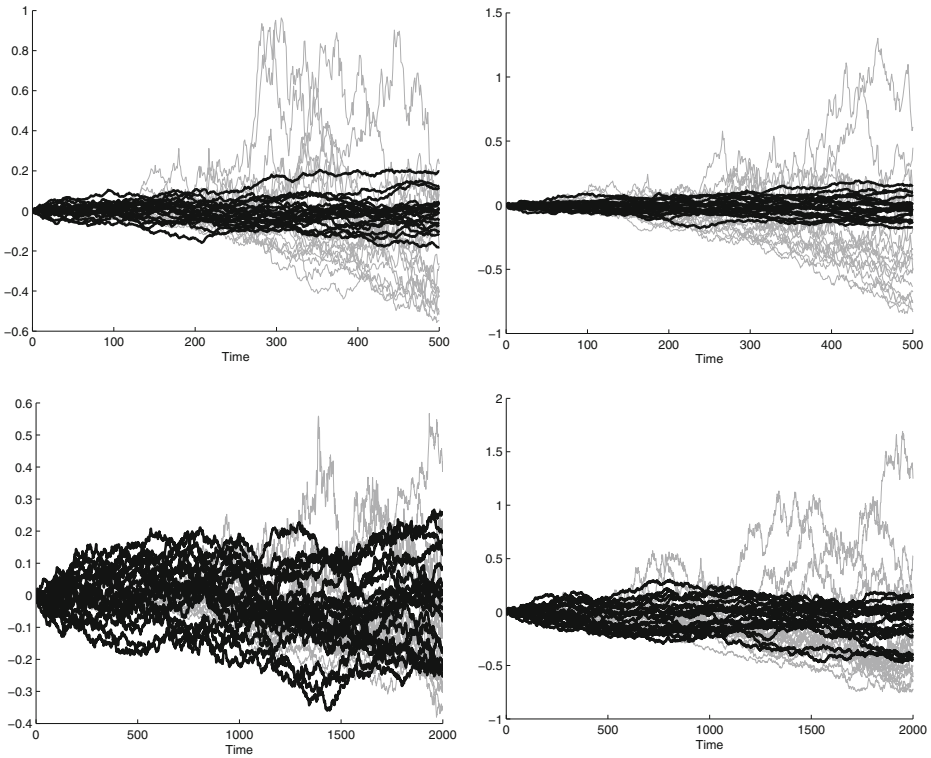

Fig. 1 $AR(1)$ example - The top two plots represent the comparison of the estimates of $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ obtained using the standard MC approach (*gray*, thin lines) and the Pairs algorithm (*black*, thick lines), where $N = 50$ for the case of equal M (*top left*) and equal cost (*top right*) respectively. The bottom two represent the comparison of the same two algorithms, but for the case, where $N = 250$ for the case of equal M (*bottom left*) and equal cost (*bottom right*)

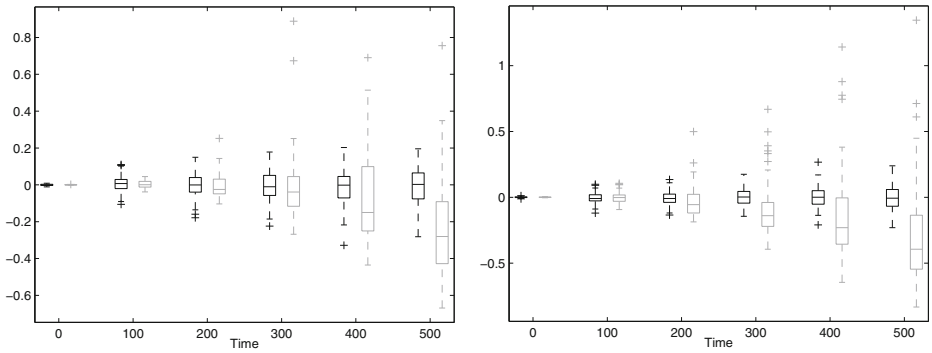

Fig. 2 $AR(1)$ example - boxplots using all of the 100 available simulated paths for the case $N = 50$ particles and equal M and equal cost. The *grey* boxplots correspond to the MC approach, and *black* ones - to the Pairs algorithm

that M and $\tilde{M}$ are chosen such that the execution times of the standard MC algorithm and the Pairs algorithm are the same. The time parameter n varies from 0 to 500 in both plots and we plot 20 independent runs of both algorithms in order to compare their variability properties.

The second row of plots in Fig. 1 consists of plots for the case of larger number of particles $N = 250$. Again, in the bottom left we are comparing the case where $M = \tilde{M} = 10^4$, and in bottom right we are comparing the equal cost case where $M = 10^4$ and $\tilde{M} = 700$. The fact that $\tilde{M}$ is lower here than in the $N = 50$ case reflects the fact that the cost of the standard MC approach is $O(\tilde{M}N)$ per time step, compared to $O(M)$ for the Pairs algorithm. We have plotted 20 independent runs for both algorithms.

Figure 2 shows boxplots based on 100 independent runs for both algorithms for the case of equal $M = \tilde{M} = 10^4$ and equal cost. We also have $N = 50$. It is apparent that the estimates of $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ that we obtain using the Pairs algorithm have much less variability than the estimates produced using the standard Monte Carlo approach with i.i.d. replicates (especially for big values of the time parameter n).

3.2 Lotka - Volterra System Example

In this section we illustrate the numerical performance of the pairs algorithm in the context of a partially observed Langevin approximation to Lotka-Volterra ODE system (Golightly and Wilkinson 2011). The signal process in the HMM is obtained from a discretization of the stochastic differential equation (SDE) $dX_t = \alpha(X_t, c)dt + \sqrt{\beta(X_t, c)}dW_t$, where $X_t = (X_{1,t}, X_{2,t})$, $W_t = (W_{1,t}, W_{,2t})$. Here W_t is a vector, each of the components of which is independent standard Brownian motion, $c = (c_1, c_2, c_3)$ are parameters and $\alpha(x, c)$ and $\beta(x, c)$ are the drift and diffusion coefficients given for the Lotka-Volterra system by

$$\alpha(x, c) = \begin{pmatrix} c_1x_1 - c_2x_1x_2 \\ c_2x_1x_2 - c_3x_2 \end{pmatrix}, \qquad \beta(x, c) = \begin{pmatrix} c_1x_1 + c_2x_1x_2 & -c_2x_1x_2 \\ -c_2x_1x_2 & c_2x_1x_2 + c_3x_2 \end{pmatrix},$$

with $x = (x_1, x_2)$.

We consider Euler discretization of the SDE with time resolution $\Delta t = 1/m$ for some $m \geqslant 1$, with the resulting process satisfying

$$X_{n+(j+1)\Delta t} - X_{n+j\Delta t} = \alpha(X_{n+j\Delta t}, c)\Delta t + \sqrt{\beta(X_{n+j\Delta t}, c)\Delta t}\chi_j \tag{15}$$

for $n \in \mathbb{N}$ and $j \in \{0, 1, \ldots, m-1\}$, where χ_j is a sequence of $\mathcal{N}(0, 1)$–independent random variables. The signal process in the HMM, denoted by $(\mathbf{X}_n)_{n\geqslant 0}$, consists of a $\mathbb{R}^2$–valued random variable $X_0 = (100, 100)$ and for $n \geqslant 1$ a $\mathbb{R}^{2m}$–valued random variable $\mathbf{X}_{n+1} = (X_{n+\Delta t}, X_{n+2\Delta t}, \ldots, X_{n+1})$. The model for the observations is $Y_n = X_n + \varepsilon_n$, where $\varepsilon_n \sim \mathcal{N}(0, \Sigma_{2\times 2})$, $\Sigma_{2\times 2} = \sigma^2 I_{2\times 2}$, where $I_{2\times 2}$ is the 2×2 identity matrix. We also assume that we have observed the process at integer times n. Following Golightly and Wilkinson (2011), we consider two values of the observation noise variance $\sigma^2 = 10$ and $\sigma^2 = 200$. We fix the rate constants $c = (c_1, c_2, c_3) = (0.5, 0.0025, 0.3)$, and we will use $m = 1$ for the discretization parameter.

We adopt the same approach to constructing the proposal kernels $(q_n)_{n\geqslant 1}$ suggested in ? [()Section 4.3]golightly2011bayesian, in which $q_n(\mathbf{x}_n, \mathbf{x}_{n+1})$ is chosen to be a tractable Gaussian approximation to the conditional density of $\mathbf{x}_{n+1}$ given $\mathbf{x}_n, \boldsymbol{y}_{n+1}$. The proposal kernel is given by

$$q_{n+1}(\mathbf{x}_n, \mathbf{x}_{n+1}) = \prod_{j=0}^{m-1} \psi_{n+(j+1)\Delta t}(x_{n+j\Delta t}, x_{n+(j+1)\Delta t})$$

where $\psi_{n+(j+1)\Delta t}(x_{n+j\Delta t}, \cdot) = \mathcal{N}(\cdot; x_{n+j\Delta t} + a_j\Delta t, b_j\Delta t)$, where $a_j = \alpha_j + \beta_j(\beta_j\Delta_j + \Sigma)^{-1}(y_{n+1} - (x_{n+j\Delta t} + \alpha_j\Delta_j))$, $b_j = \beta_j - \beta_j(\beta_j\Delta_j + \Sigma)^{-1}\beta_j\Delta t$, $\Delta_j = 1 - j\Delta t$, $\alpha_j = \alpha(x_{n+j\Delta t}, c)$, $\beta_j = \beta(x_{n+j\Delta t}, c)$. We consider the process $(\mathbf{X}_n, Y_n)_{n\geqslant 0}$ as a HMM, to which the particle algorithms are applied to.

We first obtain a reliable benchmark value of $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$, denoted by $\Xi_n^{(N,M')}$, using a single run of the Pairs algorithm with $M' = 10^6$. We compare $\Xi_n^{(N,M)}$ from the Pairs algorithm with the simple Monte Carlo approximation $\widetilde{\Xi}_n^{(N,\tilde{M})}$ based on i.i.d. replicates, defined in Eq. 13 in Fig. 3 for two different values of the observation noise - $\sigma^2 = 10$ and $\sigma^2 = 200$. In both cases we plot again $\log(\Xi_n^{(N,M)}) - \log(\Xi_n^{(N,M')})$ for the Pairs algorithm and $\log(\widetilde{\Xi}_n^{(N,\tilde{M})}) - \log(\Xi_n^{(N,M')})$ for the standard MC approach.

On the top left of Fig. 3 we have the low noise example. In this example, we set $N = 100$, $M = 10^4$ and $\tilde{M} = 300$. On the top right plot we present the large noise case where we set $N = 100$, $M = 10^5$ and $\tilde{M} = 3000$ in order to equalize the computational cost. Again, as in the previous example, we have plotted 20 independent runs for both algorithms.

In the two plots, and especially for large values of the time parameter n, the estimate that we obtain with the help of the Pairs algorithm has much less variability than the estimate calculated using standard Monte Carlo with i.i.d. replicates. We can clearly see that with the increase of the time parameter n, the rate of growth of the variability of the estimates of $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ obtained using the Pairs algorithm is far less than the corresponding rate for the standard Monte Carlo approach (using i.i.d. replicates). The observations about the variability of the estimates in Fig. 3 are also supported by the corresponding boxplots, based on 100 independent runs of the two algorithms.

Table 1 shows numerical values for Z_n^N and $\Xi_n^{(N,M)}$ for different values of the time parameter n for the Lotka–Volterra example. We see, that although the scale of the values in Table 1 is small, we still have, by Jensen's inequality, that $\mathbb{E}\left[\left(Z_n^N\right)^2\right] \geqslant \mathbb{E}\left[Z_n^N\right]^2$.

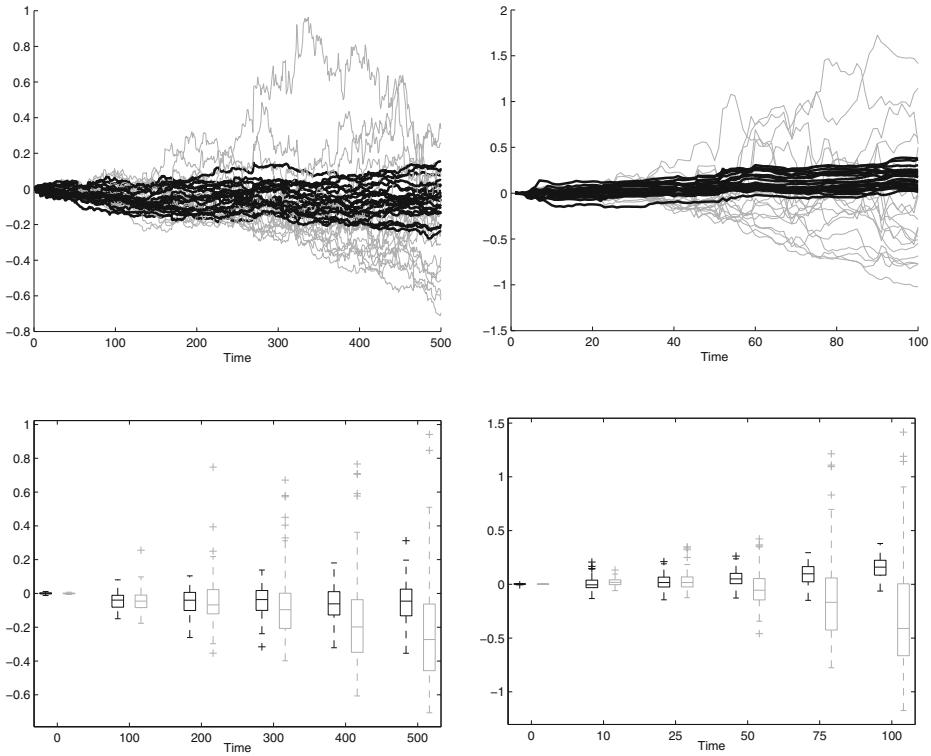

Fig. 3 Lotka - Volterra example - comparison of the estimates of $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ for the Pairs algorithm and the standard Monte Carlo approach for the case of low observation noise ($\sigma^2 = 10$, on the *left*) and large observation noise ($\sigma^2 = 200$, on the *right*). The plots are for equal time cost. Again, *grey* corresponds to the MC approach and *black* corresponds to the Pairs algorithm. The boxplots are based on 100 independent runs of the two algorithms

3.3 Estimating Monte Carlo Variance

The purpose of this example is to show that the benefits of approximating $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ using the Pairs algorithm carry over to its use within a strategy for both estimating Z_n and reporting Monte Carlo variance. As a benchmark for comparisons, we consider the following standard approach based on i.i.d. replicates of a particle filter.

MC Strategy Run $\tilde{M}$ independent particle filters, each with $\tilde{N}$ particles, to give $\left\{Z_n^{\tilde{N},j}\right\}_{j=1}^{\tilde{M}}$. Then report:

- $\widetilde{Z}_n^{(\tilde{N},\tilde{M})} = \frac{1}{\tilde{M}} \sum_{j=1}^{\tilde{M}} Z_n^{\tilde{N},j}$ as an estimate of Z_n
- $\frac{1}{\tilde{M}} \frac{1}{\tilde{M}-1} \sum_{j=1}^{\tilde{M}} \left(Z_n^{\tilde{N},j} - \widetilde{Z}_n^{(\tilde{N},\tilde{M})}\right)^2$ as an estimate of $\mathrm{Var}\left[\widetilde{Z}_n^{(\tilde{N},\tilde{M})}\right]$

The cost of this strategy is $O(\tilde{N}\tilde{M})$, and the variance estimate it delivers is a standard sample variance, thus unbiased. There are various ways that the MC strategy could be changed or augmented by using the Pairs algorithm. We consider the following:

Table 1 Estimates of Z_n and $\mathbb{E}\left[\left(Z_n^N\right)^2\right]$ (using the Pairs algorithm, hence $\Xi_n^{(N,M)}$ with $M = 10^6$) for the two cases of low and large observation noise for the Lotka-Volterra example

Time	Low noise		Large noise	
n	Z_n^N	$\Xi_n^{(N,M)}$	Z_n^N	$\Xi_n^{(N,M)}$
1	7.4×10^{-3}	4.9×10^{-5}	2.21×10^{-4}	6.3×10^{-8}
5	4.13×10^{-15}	1.6×10^{-29}	4.07×10^{-21}	3.9×10^{-41}
10	2.51×10^{-33}	4.9×10^{-66}	4.59×10^{-42}	1.15×10^{-82}
25	2.02×10^{-80}	4.9×10^{-160}	3.42×10^{-100}	2.69×10^{-198}
50	2.23×10^{-159}	8.9×10^{-318}	2.81×10^{-195}	$\leqslant 10^{-324}$
100	6.41×10^{-317}	$\leqslant 10^{-324}$	$\leqslant 10^{-324}$	$\leqslant 10^{-324}$

Pairs Strategy Run M independent particle filter algorithms, each with N particles, to give $\left\{Z_n^{N,j}\right\}_{j=1}^N$. Additionally run one instance of the Pairs algorithm with parameters (M, N), to give $\Xi_n^{(N,M)}$. Then report:

- $Z_n^{(N,M)} = \frac{1}{M}\sum_{j=1}^M Z_n^{N,j}$ as an estimate of Z_n
- $\frac{1}{M-1}\left[\Xi_n^{(N,M)} - \left(Z_n^{(N,M)}\right)^2\right]$ as an estimate of $\mathrm{Var}\left[Z_n^{(N,M)}\right]$

The cost of this strategy is $O(MN + M)$. So if for instance $N = \tilde{N}$ and $M = \tilde{M}$, the additional cost of the Pairs strategy beyond that of the MC strategy becomes negligible as N grows.

To see that the variance estimate delivered by the Pairs strategy is unbiased, note that:

$$\begin{aligned}
&\frac{M}{M-1}\mathbb{E}\left[\Xi_n^{(M,N)} - \left(Z_n^{(N,M)}\right)^2\right] \\
&= \frac{M}{M-1}\left[\mathbb{E}\left[\Xi_n^{(M,N)}\right] - \frac{1}{M^2}\sum_{j=1}^M \mathbb{E}\left[\left(Z_n^{N,j}\right)^2\right] - \frac{1}{M^2}\sum_{i\neq j}^M \mathbb{E}\left[Z_n^{N,i}\right]\mathbb{E}\left[Z_n^{N,j}\right]\right] \\
&= \frac{M}{M-1}\left[\mathbb{E}\left[\left(Z_n^N\right)^2\right] - \frac{1}{M}\mathbb{E}\left[\left(Z_n^N\right)^2\right] - \left(1 - \frac{1}{M}\right)\mathbb{E}\left[Z_n^N\right]^2\right] \\
&= \mathrm{Var}[Z_n^N] = M\mathrm{Var}\left[\frac{1}{M}\sum_{j=1}^M Z_n^{(N,j)}\right],
\end{aligned}$$

where the second equality uses the lack-of-bias property of the Pairs algorithm from Theorem 2.1, i.e. $\mathbb{E}\left[\Xi_n^{(M,N)}\right] = \mathbb{E}\left[\left(Z_n^N\right)^2\right]$.

Numerical results are shown in Fig. 4. In order to achieve better visual representation, we plot normalized estimates $Z_n^{(N,M)}/Z_n^{N'}$ and $\widetilde{Z}_n^{(\tilde{N},\tilde{M})}/Z_n^{N'}$ and their variances $\mathrm{Var}\left[Z_n^{(N,M)}\right]/\left(Z_n^{N'}\right)^2$ and $\mathrm{Var}\left[\widetilde{Z}_n^{(\tilde{N},\tilde{M})}\right]/\left(Z_n^{N'}\right)^2$, where $Z_n^{N'}$ is a reliable, benchmark estimate of Z_n obtained from a particle filter with $N' = 10^6$. We make comparisons with $N = \tilde{N} = 50$ and $M = \tilde{M} = 10^4$, with these settings in our implementation the additional cost of the Pairs strategy beyond that of the MC strategy was found to be insignificant, very similar results were obtained if the costs of the two strategies were exactly equalized.

In Fig. 4 we compare the MC and Pairs strategies. The top left shows box plots of $Z_n^{(N,M)}/Z_n^{N'}$ and $\widetilde{Z}_n^{(\tilde{N},\tilde{M})}/Z_n^{N'}$ obtained from 1000 independent realizations of the two strategies, for different values of n. The top right shows boxplots for the variance estimates, also from 1000 realizations. We can clearly see that for increasing n the estimates for the MC strategy exhibit larger variability than the estimates obtained from the Pairs strategy. On the bottom two plots of Fig. 4 we compare the kernel density estimates for $\mathrm{Var}\left[Z_n^{(N,M)}\right]/\left(Z_n^{N'}\right)^2$ and $\mathrm{Var}\left[\widetilde{Z}_n^{(\tilde{N},\tilde{M})}\right]/\left(Z_n^{N'}\right)^2$ for $n = 500$. On bottom left the estimated density is plotted, and on bottom right the log of the density is plotted, highlighting the heavier tails of the distribution for the MC strategy. The kernel density estimates in both plots were produced using a normal kernel function with bandwidths 0.06 (Pairs strategy) and 0.9 (MC strategy). The density estimates indicated a more concentrated distribution for the Pairs strategy (thick, black line) than for the MC strategy (grey line).

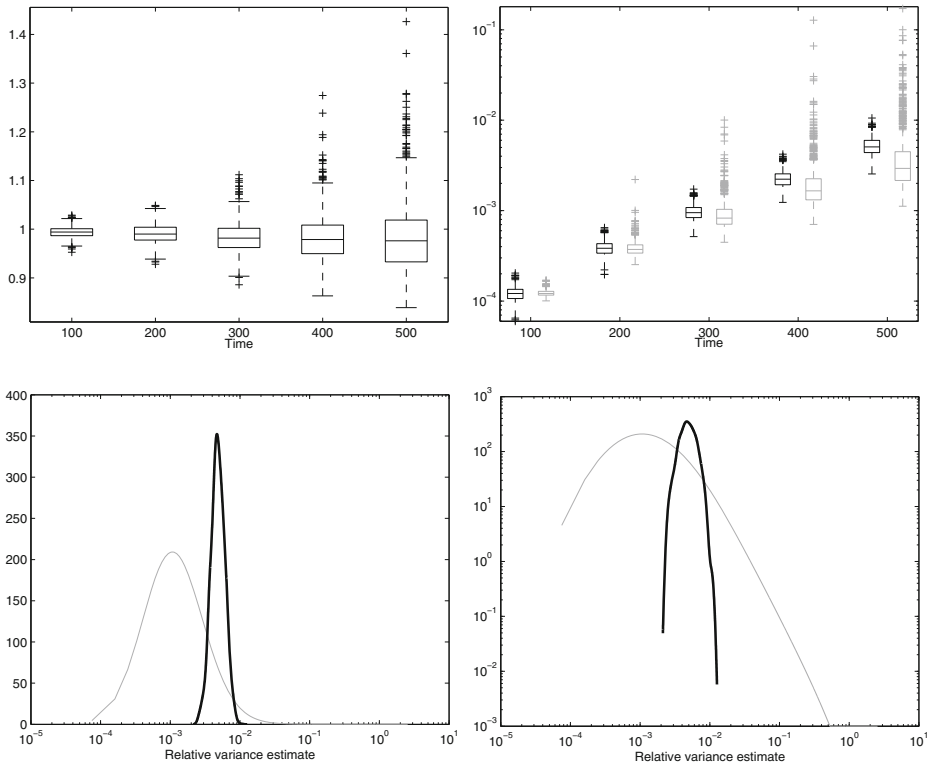

Fig. 4 $AR(1)$ example - comparison of Pairs and MC strategies for $\tilde{N} = N = 50$ particles and $\tilde{M} = M = 10^4$. On *top left* we plot the estimates of $\mathbb{E}\left[Z_n^N\right]/Z_n^{N'}$ for both MC and Pairs strategies (which are equal). On *top right* plot we compare the two strategies in terms of estimates of the relative variance $Var\left[Z_n^{(N,M)}\right]/\left(Z_n^{N'}\right)^2$ and $Var\left[\widetilde{Z}_n^{(\tilde{N},\tilde{M})}\right]/\left(Z_n^{N'}\right)^2$ respectively (the y–axis is on a log–scale). On the *bottom left* (*right*) plot we compare the kernel density estimates of the pdf (log–pdf) of the relative variance for the two strategies for time $n = 500$ (the y–axis of the bottom right plot is on a log–scale). For the bottom two plots the x–axis is on a log–scale

Acknowledgements SK would like to thank the University of Bristol for providing him with University of Bristol Postgraduate Scholarship during the time this research was done.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Auxiliary Definitions, Results and Proof of Theorem 2.1

This appendix is structured as follows. After introducing notation in Sections A.1, A.2 introduces a generic particle system, of which we show Algorithm 1 to be a special case. The account of this generic particle system and some of its properties is needed in order to derive an associated pairs particle system in Section A.3, of which we show Algorithm 2 to be a special case. The proof of Theorem 2.1, in Section A.4, rests on the key observation that the pairs particle system is also an instance of the generic particle system of Section A.2, allowing properties of the latter to be transferred to the Pairs algorithm.

A.1 Notation and Conventions

For a measurable space $(E, \mathcal{E})$, denote by $\mathcal{B}_b(E)$ the set of all $\mathbb{R}$-valued, measurable and bounded functions on E, and by $\mathcal{M}(E)$ and $\mathcal{P}(E)$ the sets of respectively measures and probability measures on $\mathcal{E}$. For $\mu \in \mathcal{M}(E)$ and $\varphi \in \mathcal{B}_b(E)$ we write $\mu(\varphi) := \int_E \varphi(x)\mu(dx)$. For a non-negative integral kernel $L : E \times \mathcal{E} \to [0, \infty)$, $\varphi \in \mathcal{B}_b(E)$ and $\mu \in \mathcal{M}(E)$, we write $L(\varphi)(x) := \int_E L(x, dy)\varphi(y)$, $(\mu L)(\cdot) := \int_E \mu(dx)L(x, \cdot)$ and for two such kernels, L and M, we write their composition as $(LM)(x, \cdot) := \int_E L(x, dx')M(x, \cdot)$. We write two-fold tensor product measures and functions as respectively $\mu^{\otimes 2} \in \mathcal{M}(E^2)$ and $\varphi \otimes \varphi \in \mathcal{B}_b(E \times E)$. For $\varphi \in \mathcal{B}_b(E \times E)$ we write the tensor product integral operator $L^{\otimes 2}(\varphi)(x, x') := \int_{E\times E} L(x, dy)L(x', dy')\varphi(y, y')$. We introduce also a measurable space $(E_0, \mathcal{E}_0)$ and use exactly similar notation when dealing with functions, measures and kernels on $(E_0, \mathcal{E}_0)$, and kernels between $(E_0, \mathcal{E}_0)$ and $(E, \mathcal{E})$.

A.2 A Generic Particle System

For each $n \geqslant 2$ let $Q_n : E \times \mathcal{E} \to (0, \infty)$ be an integral kernel such that for each $x \in E$, $Q_n(x, \cdot)$ is a finite measure on $(E, \mathcal{E})$. Then introduce

$$M_n : (x, A) \in E \times \mathcal{E} \mapsto \frac{Q_n(x, A)}{Q_n(x, E)} \in [0, 1]; \quad G_{n-1} : x \in E \mapsto Q_n(x, E) \in (0, \infty), \tag{16}$$

which are respectively a Markov kernel and a measurable, bounded, strictly positive function. Let also $Q_1 : E_0 \times \mathcal{E} \to (0, \infty)$ be a finite integral kernel, with M_1 and G_0 defined similarly to Eq. 16.

For $0 \leqslant p \leqslant n$ define $Q_{p,n} = Q_{p+1} \cdots Q_n$ with $Q_{n,n} := Id$. Fix some $\eta_0 \in \mathcal{P}(E_0)$, and define the measures $(\gamma_n)_{n\geqslant 0}$ and probability measures $(\eta_n)_{n\geqslant 1}$ by $\gamma_0 := \eta_0$ and

$$\gamma_n(\cdot) := \eta_0 Q_{0,n}(\cdot), \quad \eta_n(\cdot) := \frac{\gamma_n(\cdot)}{\gamma_n(E)}, \quad n \geqslant 1. \tag{17}$$

With these objects, and for some fixed $N \geqslant 1$, we associate a particle process $(\zeta_n)_{n\geqslant 0}$ as follows. The initial configuration $\zeta_0 = \left\{\zeta_0^i\right\}_{i=1}^N$ are independent and identically distributed

according to η_0, and the evolution of $\zeta_n = \{\zeta_n^i\}_{i=1}^N$ is described by the following probability law

$$\mathbb{P}(\zeta_n \in d\zeta_n | \zeta_0, ..., \zeta_{n-1}) := \prod_{i=1}^N \frac{\sum_{j=1}^N Q_n(\zeta_{n-1}^j, d\zeta_n^i)}{\sum_{j=1}^N Q_n(\zeta_{n-1}^j, E)} = \prod_{i=1}^N \frac{\sum_{j=1}^N G_{n-1}(\zeta_{n-1}^j) M_n(\zeta_{n-1}^j, d\zeta_n^i)}{\sum_{j=1}^N G_{n-1}(\zeta_{n-1}^j)}, \quad n \geqslant 1, \tag{18}$$

where $d\zeta_n$ is to be understood as an infinitesimal neighborhood of a point $(\zeta_n^1, ..., \zeta_n^N)$.

Let us define the empirical measures

$$\eta_n^N := N^{-1} \sum_{i=1}^N \delta_{\zeta_n^i}, \ n \geqslant 0. \tag{19}$$

$$\gamma_0^N := \eta_0^N, \quad \gamma_n^N(\cdot) := \eta_n^N(\cdot) \prod_{p=0}^{n-1} \eta_p^N(G_p), \quad n \geqslant 1. \tag{20}$$

Algorithm 1 as an Instance of the Generic Particle System

Let $(\mathsf{X}, \mathcal{X}), \pi_0, f, g$ be the ingredients of the HMM as in Section 1. To obtain Algorithm 1 as an instance of the generic particle system under the law (18), take $E_0 = \mathsf{X}$, $\mathcal{E}_0 = \mathcal{X}$, and $E = \mathsf{X}^2$, $\mathcal{E} = \mathcal{X}^{\otimes 2}$. Then for points $x = (x_1, x_2) \in E$ and $y = (y_1, y_2) \in E$, take

$$M_n(x, dy) = \delta_{x_2}(dy_1) q_n(y_1, y_2) dy_2, \quad G_{n-1}(x) = \frac{g_{n-1}(x_2) f(x_1, x_2)}{q_{n-1}(x_1, x_2)}, \quad n \geqslant 2, \tag{21}$$

and for $x \in E_0$, $y = (y_1, y_2) \in E$, take

$$M_1(x, dy) = \delta_x(dy_1) q_1(y_1, y_2) dy_2, \quad G_0(x) = \frac{g_0(x)\pi_0(x)}{q_0(x)}, \quad \eta_0 = \pi_0. \tag{22}$$

Observe then that with Z_n as in Eq. 2 and Z_n^N as in Algorithm 1,

$$\gamma_{n+1}(1) \equiv Z_n, \quad \gamma_{n+1}^N(1) \equiv Z_n^N. \tag{23}$$

Properties of the Generic Particle System

We now give a brief account of certain key properties of the particle system introduced above, which we shall later put to use in analyzing the pairs algorithm.

Remark A.1 It is known that when, for each $n \geqslant 0$,

$$\sup_x G_n(x) < \infty, \tag{24}$$

we have for any $\varphi \in \mathcal{B}_b(E)$,

$$\eta_n^N(\varphi) \xrightarrow[N \to \infty]{a.s.} \eta_n(\varphi), \qquad \gamma_n^N(\varphi) \xrightarrow[N \to \infty]{a.s.} \gamma_n(\varphi), \tag{25}$$

see e.g. (Del Moral 2004, Theorem 7.4.2). Moreover, as discussed in (Del Moral 2004, Section 9.4.1),

$$\mathbb{E}\Big[\gamma_n^N(\varphi)\Big] = \gamma_n(\varphi) = \eta_0 Q_{0,n}(\varphi), \quad \forall N \geqslant 1. \tag{26}$$

Cérou et al. (2011) have obtained second moment formulae for $\gamma_n^N(1)$ via a study of the tensor product empirical measures:

$$\left(\eta_n^N\right)^{\otimes 2} := \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\delta_{\zeta_n^i}\otimes\delta_{\zeta_n^j}$$
$$\left(\gamma_n^N\right)^{\otimes 2} := \gamma_n^N(1)^2\left(\eta_n^N\right)^{\otimes 2}.$$

Introducing the coalescence operator C which acts on bounded measurable functions F as $C(F)(x,y) = F(x,x)$, we have:

Proposition A.1 (Cérou et al. 2011, Lemma 3.2)*For any* $F \in \mathcal{B}_b(E\times E)$,

$$\mathbb{E}\left[\left(\gamma_n^N\right)^{\otimes 2}(F)\right] = \mathbb{E}\left[\eta_0^{\otimes 2}C_{\epsilon_0}Q_1^{\otimes 2}C_{\epsilon_1}\cdots Q_n^{\otimes 2}C_{\epsilon_n}(F)\right] \tag{27}$$

and in particular for $F = 1\otimes 1$,

$$\mathbb{E}\left[\gamma_n^N(1)^2\right] = \mathbb{E}\left[\eta_0^{\otimes 2}C_{\epsilon_0}Q_1^{\otimes 2}C_{\epsilon_1}\cdots C_{\epsilon_{n-1}}Q_n^{\otimes 2}(1\otimes 1)\right] \tag{28}$$

where $C_1 := C$, $C_0 := Id$ *and* $\{\epsilon_n\}_{n\geqslant 0}$ *is a sequence of i.i.d.,* $\{0,1\}$*-valued random variables with distribution*

$$\mathbb{P}(\epsilon_n = 1) = 1 - \mathbb{P}(\epsilon_n = 0) = \frac{1}{N}.$$

Proposition A.2 (Cérou et al. 2011, Corollary 1.5) *If for each* $p \geqslant 0$ *there exists a finite constant* c_p *such that*

$$\sup_{n\geqslant p}\sup_{(x,y)\in E^2}\frac{Q_{p,n}(1)(x)}{Q_{p,n}(1)(y)} \leqslant c_p, \tag{29}$$

then for any $n \geqslant 0$,

$$N > \sum_{s=0}^{n} c_s \quad\Rightarrow\quad \mathbb{E}\left[\left(\frac{\gamma_n^N(1)}{\gamma_n(1)} - 1\right)^2\right] \leqslant \frac{4}{N}\sum_{s=0}^{n} c_s.$$

Remark A.2 If for each $n \geqslant 0$

$$\delta_n := \sup_{(x,y)\in E^2}\frac{G_n(x)}{G_n(y)} < \infty \quad\text{and}\quad M_{n,n+m}(x,\cdot) \leqslant \beta_n^{(m)}M_{n,n+m}(y,\cdot), \quad \forall (x,y)\in E^2 \tag{30}$$

for some constants $m \geqslant 1$, $\beta_n^{(m)} \in [1,\infty[$, then Eq. 29 is satisfied with $c_p = \beta_p^{(m)}\prod_{p\leqslant q<p+m}\delta_q$. For a proof see e.g. (Cérou et al. 2011, Lemma 1.5). We note that the statement of (Cérou et al. 2011, Corollary 1.5) is written in terms of the condition (30), but the proof of (Cérou et al. 2011, Corollary 1.5) actually uses Eq. 29.

A.3 The Pairs Particle System

In order to derive the Pairs algorithm, our first step is to obtain in Proposition A.3 below an alternative representation of the formula on the right of Eq. 27. Define for each $n \geqslant 1$, the kernels,

$$\mathbf{Q}_n^{(N)}(x,dy) := \frac{1}{N}Q_n(\check{x},d\check{y})Q_n(\check{x},d\hat{y}) + \left(1-\frac{1}{N}\right)Q_n(\check{x},d\check{y})Q_n(\hat{x},d\hat{y}),$$

with $y = (\check{y}, \hat{y}) \in E^2$, $x = (\check{x}, \hat{x}) \in E^2$ when $n \geqslant 2$ and $x = (\check{x}, \hat{x}) \in E_0^2$ when $n = 1$. Similarly to $Q_{p,n}$ we write for $p < n$, $\mathbf{Q}_{p,n}^{(N)} := \mathbf{Q}_{p+1}^{(N)} \cdots \mathbf{Q}_n^{(N)}$ and $\mathbf{Q}_{n,n}^{(N)} := Id$. Note that we can equivalently write $\mathbf{Q}_n^{(N)}$ using the previously defined coalescence operator C as:

$$\mathbf{Q}_n^{(N)} = \frac{1}{N} C Q_n^{\otimes 2} + \left(1 - \frac{1}{N}\right) Q_n^{\otimes 2}.$$

Proposition A.3 *For any $n \geqslant 1$, $N \geqslant 2$, and $F \in \mathcal{B}_b(E \times E)$,*

$$\mathbb{E}\left[(\gamma_n^N)^{\otimes 2}(F)\right] = \eta_0^{\otimes 2} \mathbf{Q}_{0,n}^{(N)}(F_N), \tag{31}$$

where $F_N := N^{-1} C F + (1 - 1/N) F$, and in the particular case $F = 1 \otimes 1$,

$$\mathbb{E}\left[\gamma_n^N(1)^2\right] = \eta_0^{\otimes 2} \mathbf{Q}_{0,n}^{(N)}(1 \otimes 1). \tag{32}$$

Proof Starting from the identity of Proposition A.1, namely equation (27), we have

$$\begin{aligned}
&\mathbb{E}\left[(\gamma_n^N)^{\otimes 2}(F)\right] \\
&= \sum_{\epsilon_{0:n} \in \{0,1\}^{n+1}} \eta_0^{\otimes 2} C_{\epsilon_0} Q_1^{\otimes 2} C_{\epsilon_1} \cdots Q_n^{\otimes 2} C_{\epsilon_n}(F) \prod_{p=0}^{n} \left(1 - \frac{1}{N}\right)^{\mathbb{I}[\epsilon_p = 0]} \left(\frac{1}{N}\right)^{\mathbb{I}[\epsilon_p = 1]} \\
&= \sum_{\epsilon_{0:n-1} \in \{0,1\}^{n}} \int_{E_0^2 \times E^{2n}} F_N(x_n) \eta_0^{\otimes 2}(dx_0) \prod_{p=1}^{n} \left(C_{\epsilon_{p-1}} Q_p^{\otimes 2}\right)(x_{p-1}, dx_p) \\
&\quad \left(1 - \frac{1}{N}\right)^{\mathbb{I}[\epsilon_{p-1} = 0]} \left(\frac{1}{N}\right)^{\mathbb{I}[\epsilon_{p-1} = 1]} \\
&= \int_{E_0^2 \times E^{2n}} F_N(x_n) \eta_0^{\otimes 2}(dx_0) \prod_{p=1}^{n} \mathbf{Q}_p^{(N)}(x_{p-1}, dx_p) \\
&= \eta_0^{\otimes 2} \mathbf{Q}_{0,n}^{(N)}(F_N),
\end{aligned}$$

which establishes (31). For Eq. 32, note $C(1 \otimes 1) = 1 \otimes 1$ and $(\gamma_n^N)^{\otimes 2}(1 \otimes 1) = \gamma_n^N(1)^2$. □

Throughout the remainder of this section $N \geqslant 1$ is fixed. Similarly to Eq. 16, we now associate with $\left(\mathbf{Q}_n^{(N)}\right)_{)n \geqslant 1}$ collections of Markov kernels $\left(\mathbf{M}_n^{(N)}\right)_{n \geqslant 1}$ and positive functions $\left(\mathbf{G}_n^{(N)}\right)_{n \geqslant 0}$, given for $x = (\check{x}, \hat{x}) \in E^2$,

$$\mathbf{G}_{n-1}^{(N)}(x) := \mathbf{Q}_n^{(N)}(x, E \times E) = \frac{1}{N} G_{n-1}(\check{x})^2 + \left(1 - \frac{1}{N}\right) G_{n-1}(\check{x}) G_{n-1}(\hat{x}), \tag{33}$$

$$\begin{aligned}
\mathbf{M}_n^{(N)}(x, dy) &:= \frac{\mathbf{Q}_n^{(N)}(x, dy)}{\mathbf{Q}_n^{(N)}(x, E \times E)} = \frac{\mathbf{Q}_n^{(N)}(x, dy)}{\int_{E \times E} \mathbf{Q}_n^{(N)}(x, dz)} = \frac{\mathbf{Q}_n^{(N)}(x, dy)}{\mathbf{G}_{n-1}^{(N)}(x)} \\
&= p_{n-1}\left(\check{x}, \hat{x}\right) M_n\left(\check{x}, d\check{y}\right) M_n\left(\check{x}, d\hat{y}\right) \\
&\quad + (1 - p_{n-1}\left(\check{x}, \hat{x}\right)) M_n\left(\check{x}, d\check{y}\right) M_n\left(\hat{x}, d\hat{y}\right),
\end{aligned} \tag{34}$$

where

$$p_{n-1}\left(\check{x}, \hat{x}\right) := \left[1 + (N-1) \frac{G_{n-1}(\hat{x})}{G_{n-1}(\check{x})}\right]^{-1}. \tag{35}$$

Now similarly to Eq. 17, define the measures $\left(\Gamma_n^{(N)}\right)_{n\geqslant 0}$ and probability measures $\left(H_n^{(N)}\right)_{n\geqslant 1}$ according to $H_0^{(N)} := \Gamma_0^{(N)} := \eta_0^{\otimes 2}$ and

$$\Gamma_n^{(N)}(\cdot) := \eta_0^{\otimes 2}\mathbf{Q}_{0,n}^{(N)}(\cdot), \quad H_n^{(N)}(\cdot) := \frac{\Gamma_n^{(N)}(\cdot)}{\Gamma_n^{(N)}(E\times E)}, \quad n \geqslant 1. \tag{36}$$

With these objects, and for some fixed $M \geqslant 1$, we associate a particle process $(\xi_n)_{n\geqslant 0}$ as follows. The initial configuration $\xi_0 = \left\{\xi_0^1, , \ldots, \xi_0^M\right\}$ consists of M i.i.d. pairs, each $\xi_0^i = (\check{\xi}_0^i, \hat{\xi}_0^i)$ valued in E_0^2 and having distribution $H_0^{(N)} = \eta_0^{\otimes 2}$; and for $n \geqslant 1$, $\xi_n = \left\{\xi_n^1, , \ldots, \xi_n^M\right\}$ consists of M pairs, each $\xi_n^i = (\check{\xi}_n^i, \hat{\xi}_n^i)$ valued in E^2, with evolution given by:

$$\begin{aligned}\mathbb{P}(\xi_n \in d\xi_n | \xi_0, ..., \xi_{n-1}) : &= \prod_{i=1}^{M} \frac{\sum_{j=1}^{M} \mathbf{Q}_n^{(N)}(\xi_{n-1}^j, d\xi_n^i)}{\sum_{j=1}^{M} \mathbf{Q}_n^{(N)}(\xi_{n-1}^j, E)} \\ &= \prod_{i=1}^{M} \frac{\sum_{j=1}^{M} \mathbf{G}_{n-1}^{(N)}(\xi_{n-1}^j)\mathbf{M}_n^{(N)}(\xi_{n-1}^j, d\xi_n^i)}{\sum_{j=1}^{M} \mathbf{G}_{n-1}^{(N)}(\xi_{n-1}^j)}, \quad n \geqslant 1.\end{aligned} \tag{37}$$

We then introduce the empirical measures

$$H_n^{(N,M)} := M^{-1}\sum_{i=1}^{M} \delta_{\xi_n^i}, \quad n \geqslant 0, \tag{38}$$

$$\Gamma_0^{(N,M)} := H_0^{(N,M)}, \quad \Gamma_n^{(N,M)}(\cdot) := H_n^{(N,M)}(\cdot)\prod_{p=0}^{n-1} H_p^{(N,M)}(\mathbf{G}_p^{(N)}), \quad n \geqslant 1.$$

Algorithm 2 as an Instance of the pairs particle system

Let $(\mathsf{X}, \mathcal{X})$, f, g, π_0, etc. be the ingredients of the HMM, defined in Section 1. To cast Algorithm 2 as an instance of the pairs particle system described above, we just make the same choices as in Eqs. 21–22. Moreover, in that situation observe that for $\Xi_n^{(N,M)}$ as appearing in Algorithm 2,

$$\Gamma_{n+1}^{(N,M)}(1\otimes 1) \equiv \Xi_n^{(N,M)} \tag{39}$$

A.4 Proof of Theorem 2.1

To conclude the paper, we gather together various facts from the preceeding sections of the appendix and complete the proof of Theorem 2.1.

Proof of Theorem 2.1 Unless stated otherwise, throughout the proof $N \geqslant 2$ is fixed to an arbitrary value. Comparing Eq. 37 with Eq. 18, we see that the pairs particle system described in Section A.3 is itself an instance of the generic particle system described in Section A.2; in place of E_0, η_0, E, G_n, M_n etc. in the latter take E_0^2, $\eta_0^{\otimes 2}$, E^2, $\mathbf{G}_n^{(N)}$, $\mathbf{M}_n^{(N)}$ etc. This observation allows us to transfer the various properties described in Section A.2 over to the pairs particle system, as follows.

Firstly, Eqs. 24–25 read in this situation as: if for each $n \geqslant 0$,

$$\sup_x \mathbf{G}_n^{(N)}(x) < \infty, \tag{40}$$

then for any $F \in \mathcal{B}_b(E \times E)$,

$$H_n^{(N,M)}(F) \xrightarrow[M\to\infty]{a.s.} H_n^{(N)}(F), \qquad \Gamma_n^{(N,M)}(F) \xrightarrow[M\to\infty]{a.s.} \Gamma_n^{(N)}(F). \tag{41}$$

Secondly, the lack-of-bias property (26), combined with Eqs. 36 and 32, reads as:

$$\mathbb{E}\left[\Gamma_n^{(N,M)}(1\otimes 1)\right] = \Gamma_n^{(N)}(1\otimes 1) = \eta_0^{\otimes 2}\mathbf{Q}_{0,n}^{(N)}(1\otimes 1) = \mathbb{E}\left[\gamma_n^N(1)^2\right], \quad \forall M \geqslant 1. \tag{42}$$

Thirdly, Proposition A.2 reads: if for each $p \geqslant 0$ there exists a finite constant $\mathbf{c}_p$ such that

$$\sup_{n\geqslant p}\sup_{(x,y)\in E^4} \frac{\mathbf{Q}_{p,n}^{(N)}(1)(x)}{\mathbf{Q}_{p,n}^{(N)}(1)(y)} \leqslant \mathbf{c}_p, \tag{43}$$

then for any $n \geqslant 0$,

$$M > \sum_{s=0}^{n}\mathbf{c}_s \quad \Rightarrow \quad \mathbb{E}\left[\left(\frac{\Gamma_n^{(N,M)}(1\otimes 1)}{\mathbb{E}\left[\gamma_n^N(1)^2\right]} - 1\right)^2\right] \leqslant \frac{4}{M}\sum_{s=0}^{n}\mathbf{c}_s, \tag{44}$$

where in writing the l.h.s. of the inequality in Eq. 44, the identity $\Gamma_n^{(N)}(1\otimes 1) = \mathbb{E}\left[\gamma_n^N(1)^2\right]$ from Eq. 42 has been applied.

To complete the proof of Theorem 2.1 it remains to show that in the setting (21)–(22), the conditions Eqs. 9 and 10–12 imply respectively (40) and (43) for suitable constants $\mathbf{c}_p$ which do not depend on N, since then re-writting Eqs. 41, 42 and 44 using Eqs. 23 and 39 gives the claims of the Theorem.

The condition (9) does indeed imply (40), since by Eq. 33, $\sup_x \mathbf{G}_n^{(N)}(x) = \sup_x G_n(x)^2$ for any N. It remains to establish (43). We first observe that with G_p as in Eqs. 21–22, conditions (10)–(11) imply that there for each $p \geqslant 0$,

$$\mathbf{d}_p := \sup_{x,y}\frac{\mathbf{G}_p^{(N)}(x)}{\mathbf{G}_p^{(N)}(y)} = \sup_{x,y}\frac{G_p(x)^2}{G_p(y)^2} \leqslant \left(\frac{w_p^+}{w_p^-}\right)^2 < +\infty.$$

Now consider (43) for some given p. When $n \leqslant p+1$,

$$\frac{\mathbf{Q}_{p,n}^{(N)}(1)(x)}{\mathbf{Q}_{p,n}^{(N)}(1)(y)} \leqslant \mathbf{d}_p.$$

For $n \geqslant p+2$, suppose there exist contants $0 < \mathbf{k}_p^- \leqslant \mathbf{k}_p^+ < +\infty$ independent of N, and $\mathbf{m}_p^{(N)} \in \mathcal{P}(\mathsf{X}^2 \times \mathsf{X}^2)$ such that

$$\mathbf{k}_p^-\mathbf{m}_p^{(N)}(\cdot) \leqslant \mathbf{Q}_{p,p+2}^{(N)}(x,\cdot) \leqslant \mathbf{k}_p^+\mathbf{m}_p^{(N)}(\cdot), \quad \forall x. \tag{45}$$

Then

$$\frac{\mathbf{Q}_{p,n}^{(N)}(1)(x)}{\mathbf{Q}_{p,n}^{(N)}(1)(y)} = \frac{\mathbf{Q}_{p,p+2}^{(N)}\mathbf{Q}_{p+2,n}^{(N)}(1)(x)}{\mathbf{Q}_{p,p+2}^{(N)}\mathbf{Q}_{p+2,n}^{(N)}(1)(y)} \leqslant \frac{\mathbf{k}_p^+}{\mathbf{k}_p^-}\frac{\mathbf{m}_p^{(N)}\mathbf{Q}_{p+2,n}^{(N)}(1)}{\mathbf{m}_p^{(N)}\mathbf{Q}_{p+2,n}^{(N)}(1)} = \frac{\mathbf{k}_p^+}{\mathbf{k}_p^-},$$

and Eq. 43 would then hold with $\mathbf{c}_p := \mathbf{d}_p \vee \frac{\mathbf{k}_p^+}{\mathbf{k}_p^-}$. Thus to complete the proof we shall show that conditions (10)–(12) imply (45). To this end note that:

$$\begin{aligned}\mathbf{Q}_{p,p+2}^{(N)} &= \left(1-\frac{1}{N}\right)\left[\frac{1}{N}C+\left(1-\frac{1}{N}\right)Id\right]Q_{p+1}^{\otimes 2}Q_{p+2}^{\otimes 2}\\ &\quad+\frac{1}{N}\left[\frac{1}{N}C+\left(1-\frac{1}{N}\right)Id\right]Q_{p+1}^{\otimes 2}CQ_{p+2}^{\otimes 2},\end{aligned} \tag{46}$$

and with

$$\mathbf{k}_p^- := \left(w_p^- w_{p+1}^- \epsilon_{p+1}^-\right)^2, \qquad \mathbf{k}_p^+ := \left(w_p^+ w_{p+1}^+ \epsilon_{p+1}^+\right)^2,$$

for all $x = (x_1, x_2)$,

$$\mathbf{k}_p^- \mu_{p+1}^{\otimes 2}(dy_1)q_{p+2}^{\otimes 2}(y_1, dy_2) \leqslant Q_{p+1}^{\otimes 2}Q_{p+2}^{\otimes 2}(x, dy) \leqslant \mathbf{k}_p^+ \mu_{p+1}^{\otimes 2}(dy_1)q_{p+2}^{\otimes 2}(y_1, dy_2) \tag{47}$$

and

$$\begin{aligned}\mathbf{k}_p^- \int_{\mathsf{X}} \mu_{p+1}(dz)\delta_z^{\otimes 2}(dy_1)q_{p+2}^{\otimes 2}(y_1, dy_2) &\leqslant Q_{p+1}^{\otimes 2}CQ_{p+2}^{\otimes 2}(x, dy) \\ &\leqslant \mathbf{k}_p^+ \int_{\mathsf{X}} \mu_{p+1}(dz)\delta_z^{\otimes 2}(dy_1)q_{p+2}^{\otimes 2}(y_1, dy_2),\end{aligned} \tag{48}$$

where $\delta_z(\cdot)$ is the Dirac measure on X located at z, and $dy = dy_1 dy_2$ is to be understood as the infinitesimal neighbourhood of $y = (y_1, y_2) \in \mathsf{X}^2 \times \mathsf{X}^2$. Combining (46)–(48) we find that (45) holds with

$$\mathbf{m}_p^{(N)}(dy) := \left(1-\frac{1}{N}\right)\mu_{p+1}^{\otimes 2}(dy_1)q_{p+2}^{\otimes 2}(y_1, dy_2) + \frac{1}{N}\int_{\mathsf{X}} \mu_{p+1}(dz)\delta_z^{\otimes 2}(dy_1)q_{p+2}^{\otimes 2}(y_1, dy_2).$$

□

References

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

Bhadra A, Ionides EL (2014) Adaptive particle allocation in iterated sequential Monte Carlo via approximating meta-models. Stat Comput 1–15

Cérou F, Del Moral P, Guyader A (2011) A nonasymptotic theorem for unnormalized Feynman-Kac particle models. Annales de L'Institut Henri Poincare Section Physique Theorique 47:629–649. doi:10.1214/10-AIHP358

Chan HP, Lai TL (2013) A general theory of particle filters in hidden Markov models and some applications. Ann Stat 41(6):2877–2904

Del Moral P (2004) Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications. Probability and its Applications. Springer Verlag, New York

Doucet A, de Freitas N, Gordon N (2001) Sequential Monte Carlo Methods in Practice (Information Science and Statistics). Springer-Verlag New York, Inc.

Golightly A, Wilkinson DJ (2011) Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. Interface Focus 1(6):807–820

Golightly A, Henderson DA, Sherlock C (2015) Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. Stat Comput 25(5):1039–1055

Lee A, Whiteley N (2015) Variance estimation and allocation in the particle filter. arXiv:150900394

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