Ann Inst Stat Math (2016) 68:1025–1053
DOI 10.1007/s10463-015-0524-y

Antithetic sampling for sequential Monte Carlo methods with application to state-space models

Svetlana Bizjajeva[1] · **Jimmy Olsson**[2]

Received: 1 October 2013 / Revised: 1 July 2014 / Published online: 14 July 2015
© The Institute of Statistical Mathematics, Tokyo 2015

Abstract In this paper, we cast the idea of antithetic sampling, widely used in standard Monte Carlo simulation, into the framework of sequential Monte Carlo methods. We propose a version of the standard auxiliary particle filter where the particles are mutated blockwise in such a way that all particles within each block are, first, offspring of a common ancestor and, second, negatively correlated conditionally on this ancestor. By deriving and examining the weak limit of a central limit theorem describing the convergence of the algorithm, we conclude that the asymptotic variance of the produced Monte Carlo estimates can be straightforwardly decreased by means of antithetic techniques when the particle filter is close to fully adapted, which involves approximation of the so-called optimal proposal kernel. As an illustration, we apply the method to optimal filtering in state-space models.

Keywords Antithetic sampling · Central limit theorem · Optimal filtering · Optimal kernel · Particle filter · State-space models

This work is supported by the Swedish Research Council, Grant 2011-5577.

✉ Jimmy Olsson
jimmyol@kth.se

Svetlana Bizjajeva
svetik@maths.lth.se

[1] Centre for Mathematical Sciences, Lund University, Box 118, 221 00 Lund, Sweden

[2] Department of Mathematics, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

1 Introduction

Sequential Monte Carlo (SMC) *methods*—alternatively termed *particle filters*—refer to a collection of algorithms which approximate recursively a sequence of target measures (referred typically to as the *Feynman–Kac flow*) by a sequence of empirical distributions associated with properly weighted samples of *particles*. These methods have received a lot of attention during the last decades and are at present applied within a wide range of scientific disciplines; see Künch (2013), Fearnhead (1998), and Doucet et al. (2001) for introductions. Doucet and Johansen (2011) provide a survey of recent developments of the SMC methodology and comprehensive treatments of theoretical aspects of SMC algorithms are given by Del Moral (2004) and Cappé et al. (2005).

In standard SMC methods, two main operations are alternated: in the *mutation step*, the particles are propagated according to a Markovian kernel and associated with importance sampling weights proportional to the Radon–Nikodym derivative of the target measure with respect to the instrumental distribution of the particles. In the subsequent *selection step*, the particle sample is transformed by selecting new particles from the current (mutated) ones using the normalised importance weights as probabilities of selection. This step serves to eliminate or duplicate particles with small or large weights, respectively.

In this paper, we propose a modification of the *auxiliary particle filter* (APF) (introduced originally by Pitt and Shephard 1999) that relies on the classical idea of *antithetic sampling* used in standard Monte Carlo estimation. When estimating some expectation

$$I(f) \triangleq \int_{\mathbb{R}} f(x)p(x)\,\mathrm{d}x,$$

where p is a probability density function and f is some given real-valued target function, the unbiased estimator

$$\hat{I}^N(f) \triangleq \frac{1}{2N}\sum_{i=1}^{N}\left[f(\xi_i) + f(\xi_i')\right]$$

of $I(f)$, where both $\{\xi_i\}_{i=1}^N$ and $\{\xi_i'\}_{i=1}^N$ are samples from p, is more efficient (has lower variance) than the standard Monte Carlo estimator based on a sample of $2N$ independent and identically distributed draws if each pair of variables $f(\xi_i)$ and $f(\xi_i')$ are *negatively correlated* for all $i \in \{1, \ldots, N\}$. In this setting, the variables $f(\xi_i)$ and $f(\xi_i')$ are referred to as *antithetic variables*. Antithetically coupled variables can be generated in different ways, and in Sect. 2, we discuss how this can be achieved by means of the well-known *permuted displacement method* (Arvidsen and Johnsson 1982). To allow for antithetic acceleration within the SMC framework, we introduce (in Sect. 2) a version of the standard APF where the particles are mutated *blockwise* in such a way that all particles within each block are, first, offspring of a common ancestor and, second, statistically dependent conditionally on this ancestor. Moreover, in Sect. 3, we establish convergence results for our proposed method in the sense of convergence in probability and weak convergence. By examining the weak limit of the obtained central limit theorem (CLT) in Corollary 2, we conclude that the asymptotic variance of the produced Monte Carlo estimates is decreased when the particle filter is close to *fully adapted* (in which case close to uniform importance weights are obtained by means of approximation of the so-called *optimal kernel*; see Pitt and Shephard 1999) and the inherent correlation structure of each block is negative. Finally, in the implementation part, Sect. 4, we apply our algorithm to *optimal filtering* in *state-space models* and benchmark its performance on a nosily observed ARCH model as well as a univariate growth model. The outcome of the simulations indicates that introducing antithetically coupled particles provides, for these models and compared to the standard APF, a significant gain of accuracy at a lowered computational cost.

1.1 Notation and definitions

To state precisely our results and keep the presentation streamlined, we preface the description of the algorithm with some measure-theoretic notation. In the following, we assume that all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A state space Ξ is called *general* if it is equipped with a countably generated σ-field $\mathcal{B}(\Xi)$, and we denote by $\mathcal{M}(\Xi)$ and $\mathcal{F}(\Xi)$ the sets of measures on $(\Xi, \mathcal{B}(\Xi))$ and real-valued $\mathcal{B}(\Xi)$-measurable functions, respectively. For any measure $\mu \in \mathcal{M}(\Xi)$ and function $f \in \mathcal{F}(\Xi)$ satisfying $\int_\Xi |f(\xi)|\, \mu(\mathrm{d}\xi) < \infty$, we denote $\mu(f) \triangleq \int_\Xi f(\xi)\, \mu(\mathrm{d}\xi)$. A transition kernel K from $(\Xi, \mathcal{B}(\Xi))$ to some other state space $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ is called *finite* if $K(\xi, \tilde{\Xi}) < \infty$ for all $\xi \in \Xi$ and *Markovian* if, in addition, $K(\xi, \tilde{\Xi}) = 1$ for all $\xi \in \Xi$. A finite transition kernel K induces two operators, the first transforming a $\mathcal{B}(\Xi) \otimes \mathcal{B}(\tilde{\Xi})$-measurable function f satisfying $\int_{\tilde{\Xi}} |f(\xi, \tilde{\xi})|\, K(\xi, \mathrm{d}\tilde{\xi}) < \infty$ into the function

$$K(\cdot, f) : \Xi \ni \xi \mapsto \int_{\tilde{\Xi}} f(\xi, \tilde{\xi})\, K(\xi, \mathrm{d}\tilde{\xi}) \tag{1}$$

in $\mathcal{F}(\Xi)$ (here $\mathcal{B}(\Xi) \otimes \mathcal{B}(\tilde{\Xi})$ denotes the product σ-field); the other transforms any measure $\nu \in \mathcal{M}(\Xi)$ into the measure

$$\nu K(\cdot) : \mathcal{B}(\tilde{\Xi}) \ni \mathsf{A} \mapsto \int_\Xi K(\xi, \mathsf{A})\, \nu(\mathrm{d}\xi) \tag{2}$$

in $\mathcal{M}(\tilde{\Xi})$. Finally, to describe lucidly joint distributions associated with Markovian transitions, we define the *outer product*, denoted by $K \otimes T$, of a kernel K from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ and a kernel T from $(\Xi \times \tilde{\Xi}, \mathcal{B}(\Xi) \otimes \mathcal{B}(\tilde{\Xi}))$ to some other state space $(\bar{\Xi}, \mathcal{B}(\bar{\Xi}))$ as the kernel

$$K \otimes T(\xi, \mathsf{A}) \triangleq \iint_{\tilde{\Xi} \times \bar{\Xi}} \mathbb{1}_\mathsf{A}(\tilde{\xi}, \bar{\xi})\, K(\xi, \mathrm{d}\tilde{\xi})\, T(\xi, \tilde{\xi}, \mathrm{d}\bar{\xi}),$$
$$\xi \in \Xi, \mathsf{A} \in \mathcal{B}(\tilde{\Xi}) \otimes \mathcal{B}(\bar{\Xi}), \tag{3}$$

from $(\Xi, \mathcal{B}(\Xi))$ to the product space $(\tilde{\Xi} \times \bar{\Xi}, \mathcal{B}(\tilde{\Xi}) \otimes \mathcal{B}(\bar{\Xi}))$.

1.2 List of notation

Notation	Definition
α	Block size
AN	Asymptotically normal
APF	Auxiliary particle filter
$\mathbb{C}_{m,n}$	(14)
EA	Extreme antithesis
$\mathcal{F}_{N,\ell}$	(7)
$\mathcal{F}(\Xi)$	The space of measurable functions on Ξ
Φ_k	(9)
ϕ_n	(26)
g_n	(25)
$\bar{\gamma}[\Psi]$	(20)
$\tilde{\gamma}$	(16)
$K(\cdot, f)$	(1)
L	(5)
$\mathcal{L}$	(11)
$\mathbb{M}_{m,n}$	(13)
μ	(5)
νK	(2)
$\otimes$	(3)
π_n	(37)
PNA	Pairwise negatively associated
$\psi_{N,i}$	Adjustment multiplier (see Algorithm 2)
Q	(25)
R_k	(6)
$\mathcal{R}_{m,k}$	(10)
Ω_N	(4)
$\tilde{\omega}_{N,\alpha(j-1)+k}$	(8)
$\bar{\sigma}^2[\Psi]$	(19)
$\tilde{\sigma}^2$	(17)
$(\Xi, \mathcal{B}(\Xi))$	General state space

2 Auxiliary particle filter with blockwise correlated mutation

In the following, we say that a collection of random variables (particles) $\{\xi_{N,i}\}_{i=1}^{M_N}$, taking values in some state space Ξ, and associated nonnegative weights $\{\omega_{N,i}\}_{i=1}^{M_N}$ *targets* a probability measure $\nu \in \mathcal{M}(\Xi)$ if

$$\Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f(\xi_{N,i}) \backsimeq \nu(f),$$

with

$$\Omega_N \triangleq \sum_{i=1}^{M_N} \omega_{N,i}, \tag{4}$$

for all functions f in some specified subset of $\mathcal{F}(\Xi)$. Here $\{M_N\}_{N=0}^{\infty}$ is an increasing sequence of integers. The set $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ is referred to as a *weighted sample* on Ξ. In this paper, we study the problem of transforming a weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ targeting $\nu \in \mathcal{M}(\Xi)$ into a weighted sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha M_N}$, $\alpha \in \mathbb{N}^*$, targeting the probability measure

$$\mu(\mathsf{A}) = \frac{\nu L(\mathsf{A})}{\nu L(\tilde{\Xi})}, \quad \mathsf{A} \in \mathcal{B}(\tilde{\Xi}), \tag{5}$$

where L is some (possibly unnormalised) finite transition kernel from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. Feynman–Kac transitions of type (5) occur within a variety of fields (see Del Moral 2004, for examples in, e.g. quantum physics and biology) and in Sect. 4 we show in detail how the flow of posterior distributions of the noisily observed Markov chain (state signal) of a state-space model can be generated according to (5). The transformation is carried out by, first, drawing particle positions $\{\tilde{\xi}_{N,i}\}_{i=1}^{\alpha M_N}$ according to, for $j \in \{1, \ldots, M_N\}$, $k \in \{1, \ldots, \alpha\}$, and $\mathsf{A} \in \mathcal{B}(\tilde{\Xi})$,

$$\begin{aligned}
&\mathbb{P}\left(\tilde{\xi}_{N,\alpha(j-1)+k} \in \mathsf{A} \mid \mathcal{F}_{N,\alpha(j-1)+k-1}\right) \\
&\quad = R_k\left(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+1}, \ldots, \tilde{\xi}_{N,\alpha(j-1)+k-1}, \mathsf{A}\right),
\end{aligned} \tag{6}$$

where we have defined the σ-fields

$$\mathcal{F}_{N,\ell} \triangleq \sigma\left(\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}, \{\tilde{\xi}_{N,j}\}_{j=1}^{\ell}\right), \quad \ell \in \{0, \ldots, \alpha M_N\}, \tag{7}$$

and each R_k is a Markovian kernel from $(\Xi \times \tilde{\Xi}^{k-1}, \mathcal{B}(\Xi) \otimes \mathcal{B}(\tilde{\Xi})^{\otimes(k-1)})$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ such that $L(\xi, \cdot) \ll R_k(\xi, \cdot)$ for all $\xi \in \Xi$. Hence, using the kernel outer product notation $\otimes$ defined in (3), the joint distribution, conditional on $\mathcal{F}_{N,\alpha(j-1)}$, of each block $\{\tilde{\xi}_{N,\alpha(j-1)+k}\}_{k=1}^{\alpha}$ can be expressed as $\otimes_{k=1}^{\alpha} R_k(\xi_{N,j}, \cdot)$. Second, these particles are associated with the weights

$$\tilde{\omega}_{N,\alpha(j-1)+k} = \omega_{N,j} \Phi_k(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+k}) \tag{8}$$

with

$$\Phi_k(\xi, \tilde{\xi}) \triangleq \frac{\mathrm{d}L(\xi, \cdot)}{\mathrm{d}\mathcal{R}_{0,k}(\xi, \cdot)}(\tilde{\xi}), \quad (\xi, \tilde{\xi}) \in \Xi \times \tilde{\Xi}, \tag{9}$$

and, for integers $0 \le m < k$ and $\mathsf{A} \in \mathcal{B}(\tilde{\Xi})$,

$$\begin{aligned}
\mathcal{R}_{m,k}(\xi, \tilde{\xi}_{1:m}, \mathsf{A}) &\triangleq \otimes_{i=m+1}^{k} R_i(\xi, \tilde{\xi}_{1:m}, \tilde{\Xi}^{k-m-1} \times \mathsf{A}) \\
&= \int_{\tilde{\Xi}} \cdots \int_{\tilde{\Xi}} R_k(\xi, \tilde{\xi}_{1:k-1}, \mathsf{A}) \prod_{\ell=m+1}^{k-1} R_\ell(\xi, \tilde{\xi}_{1:\ell-1}, \mathrm{d}\tilde{\xi}_\ell),
\end{aligned} \tag{10}$$

Algorithm 1 Blockwise correlated mutation

Require: $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ targets ν.
1: **for** $j \leftarrow 1, \dots, M_N$ **do**
2: draw $\{\tilde{\xi}_{N,\alpha(j-1)+k}\}_{k=1}^{\alpha} \sim \bigotimes_{k=1}^{\alpha} R_k(\xi_{N,j}, \cdot)$,
3: set, for $k \in \{1, \dots \alpha\}$,

$$\tilde{\omega}_{N,\alpha(j-1)+k} \leftarrow \omega_{N,j} \Phi_k(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+k}),$$

4: **end for**
5: let $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha M_N}$ approximate μ.

Algorithm 2 APF with blockwise correlated mutation

Require: $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ targets ν.
1: draw $\{I_{N,j}\}_{j=1}^{\tilde{M}_N} \sim \mathsf{Mult}(\tilde{M}_N, \{\omega_{N,i}\psi_{N,i} / \sum_{\ell=1}^{M_N} \omega_{N,\ell}\psi_{N,\ell}\}_{i=1}^{M_N})$,
2: **for** $j \leftarrow 1, \dots, \tilde{M}_N$ **do**
3: draw $\{\tilde{\xi}_{N,\alpha(j-1)+k}\}_{k=1}^{\alpha} \sim \otimes_{k=1}^{\alpha} R_k(\xi_{N,I_{N,j}}, \cdot)$,
4: set, for $k \in \{1, \dots \alpha\}$,

$$\tilde{\omega}_{N,\alpha(j-1)+k} \leftarrow \psi_{N,I_{N,j}}^{-1} \Phi_k(\xi_{N,I_{N,j}}, \tilde{\xi}_{N,\alpha(j-1)+k}),$$

5: **end for**
6: let $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha \tilde{M}_N}$ approximate μ.

where we have introduced vector notation $a_{m:n} \triangleq (a_m, a_{m+1}, \dots, a_n)$ with the convention $a_{m:n} = \varnothing$ if $m > n$. Thus $\mathcal{R}_{m,k}(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+1:\alpha(j-1)+m}, \cdot)$ is the distribution of $\tilde{\xi}_{N,\alpha(j-1)+k}$ conditionally on $\mathcal{F}_{N,\alpha(j-1)+m}$. Finally, we take $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha M_N}$ as an approximation of μ. This *blockwise mutation operation*, which extends, since it allows for statistically dependent particles within each block, the blockwise mutation operation suggested by Douc and Moulines (2008), is summarised in Algorithm 1.

2.1 Blockwise correlated mutation with resampling

In the sequential context, where the problem consists in estimating a *sequence* of measures generated recursively according to mappings of form (5) (see Sect. 4), it is, in order to avoid *weight degeneracy* (see, e.g. Cappé et al. 2005, Section 7.3.1, for a discussion), essential to combine the correlated blockwise mutation operation described in Algorithm 1 with a prefatory *resampling operation* where particles having small weights are eliminated and those having large ones are duplicated. As observed by Pitt and Shephard (1999) (see also Douc et al. 2008 for a theoretical study), the variance of the produced SMC estimates can be reduced efficiently by introducing, as in the APF, a set $\{\psi_{N,i}\}_{i=1}^{M_N}$ of *adjustment multiplier weights* and selecting the particles with probabilities proportional to $\{\omega_{N,i}\psi_{N,i}\}_{i=1}^{M_N}$. This gives us the scheme described in Algorithm 2, where Mult denotes the multinomial distribution.

2.2 Antithetic blockwise mutation with resampling

The main motivation of Pitt and Shephard (1999) for introducing the adjustment multiplier weights was the possibility of designing these in such a manner that the resulting (second stage) particle weights $\{\tilde{\omega}_{N,i}\}_{i=1}^{\alpha \tilde{M}_N}$ become close to uniform; in this case, in which the APF is referred to as *fully adapted*, the instrumental and target distributions of the APF coincide. Adapting fully the APF involves typically some approximation of the so-called *optimal* proposal kernel $L(\xi, \cdot)/L(\xi, \tilde{\Xi})$. Indeed, let $\mathcal{L}$ be a kernel from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ such that

$$\mathcal{L}(\xi, \mathsf{A}) \approx L(\xi, \mathsf{A}), \quad (\xi, \mathsf{A}) \in \Xi \times \mathcal{B}(\tilde{\Xi}); \tag{11}$$

then Algorithm 2 with $\psi_{N,i} = \mathcal{L}(\xi_{N,i}, \tilde{\Xi})$ and $\mathcal{R}_{0,k}(\xi, \cdot) = \mathcal{L}(\xi, \cdot)/\mathcal{L}(\xi, \tilde{\Xi})$ for all $i \in \{1, \ldots, M_N\}$ and $k \in \{1, \ldots, \alpha\}$ returns a close to uniformly weighted particle sample, since with this parametrisation,

$$\begin{aligned}\tilde{\omega}_{N,\alpha(j-1)+k} &= \mathcal{L}^{-1}(\xi_{N,I_{N,j}}, \tilde{\Xi}) \frac{\mathrm{d}L(\xi_{N,I_{N,j}}, \cdot)}{\mathrm{d}\mathcal{R}_{0,k}(\xi_{N,I_{N,j}}, \cdot)}(\tilde{\xi}_{N,\alpha(j-1)+k}) \\ &= \frac{\mathrm{d}L(\xi_{N,I_{N,j}}, \cdot)}{\mathrm{d}\mathcal{L}(\xi_{N,I_{N,j}}, \cdot)}(\tilde{\xi}_{N,\alpha(j-1)+k}) \\ &\approx 1.\end{aligned}$$

Considering optimal filtering in state-space models, full adaptation can be achieved in the case where the observation equation is linear/Gaussian and the state equation is possibly nonlinear with additive Gaussian noise (see Cappé et al. 2005, Section 7.2.2.2, for details). Moreover, as shown in Sect. 4, full adaptation is straightforward for particle approximation of the so-called *predictor distributions* in state-space models, as the fully adapted algorithm corresponds to the standard *bootstrap particle filter* in this case. In the general case, methods for approximating the optimal kernel have been proposed by several authors; see e.g. Pitt and Shephard (1999) and Doucet et al. (2000), and Cornebise et al. (2014).

For our purposes, putting the APF in a close to fully adapted mode is attractive from another point of view: the close to uniform weights render efficient antithetic acceleration of the standard APF possible, which might reduce significantly the variance of the produced SMC estimates. Hence, the aim of this paper is to justify, in theory as well as in simulations, Algorithm 3, describing a particular parametrisation of Algorithm 2 in which $\mathcal{L}$ and f denote a given approximation of L and some given objective function, respectively.

Step (7) in Algorithm 3 can be carried out in several different ways, and we refer to Sect. 4 for practical implementations. The simplest way to introduce negative correlation between two real-valued random variables is to use a pair (U, U') of uniforms, where $U = r$, $U' = 1 - r$, and $r \sim \mathcal{U}(0, 1)$ is uniformly distributed (on $(0, 1)$). Such a coupling has the *extreme antithesis* (EA) *property*: if F is an arbitrary distribution function on $\mathbb{R}$, then the correlation between $\xi = F^{\leftarrow}(U)$ and $\xi' = F^{\leftarrow}(U')$, $F^{\leftarrow}$ denoting the inverse of F, achieves the minimal possible value subject to the

Algorithm 3 APF with antithetic blockwise mutation

Require: $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ targets ν.
1: **for** $i \leftarrow 1, \ldots, M_N$ **do**
2: compute an approximation $\mathcal{L}(\xi_{N,i}, \cdot)$ of $L(\xi_{N,i}, \cdot)$ satisfying $L(\xi_{N,i}, \cdot) \ll \mathcal{L}(\xi_{N,i}, \cdot)$,
3: set $\psi_{N,i} \leftarrow \mathcal{L}(\xi_{N,i}, \tilde{\Xi})$,
4: **end for**
5: draw $\{I_{N,j}\}_{j=1}^{\tilde{M}_N} \sim \mathsf{Mult}(\tilde{M}_N, \{\omega_{N,i}\psi_{N,i} / \sum_{\ell=1}^{M_N} \omega_{N,\ell}\psi_{N,\ell}\}_{i=1}^{M_N})$,
6: **for** $j \leftarrow 1, \ldots, \tilde{M}_N$ **do**
7: simulate, using an appropriate family of kernels $\{R_k\}_{k=1}^{\alpha}$,

$$\{\tilde{\xi}_{N,\alpha(j-1)+k}\}_{k=1}^{\alpha} \sim \otimes_{k=1}^{\alpha} R_k(\xi_{N,I_{N,j}}, \cdot)$$

in such a manner that $\mathcal{R}_{0,k}(\xi_{N,I_{N,j}}, \cdot) = \mathcal{L}(\xi_{N,I_{N,j}}, \cdot)/\mathcal{L}(\xi_{N,I_{N,j}}, \tilde{\Xi})$ and the real-valued variables $\{f(\tilde{\xi}_{N,\alpha(j-1)+k})\}_{k=1}^{\alpha}$ are, conditionally on $\xi_{N,I_{N,j}}$, mutually negatively correlated,
8: **for** $k \leftarrow 1, \ldots, \alpha$ **do**

$$\tilde{\omega}_{N,\alpha(j-1)+k} \leftarrow \frac{\mathrm{d}L(\xi_{N,I_{N,j}}, \cdot)}{\mathrm{d}\mathcal{L}(\xi_{N,I_{N,j}}, \cdot)}(\tilde{\xi}_{N,\alpha(j-1)+k})$$

9: **end for**
10: **end for**
11: let $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha\tilde{M}_N}$ approximate μ.

constraint that both ξ and ξ' are distributed according to F. Since (U, U') achieves EA simultaneously for all F, this implies immediately that the strategy achieves EA also for variates $g(\xi)$ and $g(\xi')$, where $g : \mathbb{R} \to \mathbb{R}$ is any monotone function such that $\int g^2(\xi)\, F(\mathrm{d}\xi) < \infty$. This remarkable observation is related to the fact that the construction (U, U') satisfies the stronger property of *negative association*, which requires that the negative correlation is preserved by monotone transformations. The following definition, adopted form Craiu and Meng (2005), extends this property to an arbitrary number of variates.

Definition 1 (*pairwise negative association*) The random variables $\{\xi_i\}_{i=1}^n$ are said to be *pairwise negatively associated* (PNA) if, for any nondecreasing (or non-increasing) functions f_1, f_2 and $(i, j) \in \{1, \ldots, n\}^2$ such that $i \neq j$,

$$\mathrm{Cov}[f_1(\xi_i), f_2(\xi_j)] \leq 0$$

whenever this covariance is well defined.

In the light of the previous it is appealing to mutate, in Step (7) in Algorithm 3, the particles in such a way that the α offspring particles of a certain block are conditionally EA given the common ancestor. A rather generic way to achieve this goes via the *permuted displacement method* (developed by Arvidsen and Johnsson 1982) presented below, where S_α denotes the set of all possible permutations of the numbers $\{1, \ldots, \alpha\}$.

In this setting, Craiu and Meng (2005, Theorem 3) showed that the uniformly distributed variates $\{U_i\}_{i=1}^{\alpha}$ produced in Algorithm 4 are PNA for $\alpha \leq 3$. For $\alpha \geq 4$, one has not at present been able to neither prove nor refute a similar result. Thus, Step

Algorithm 4 Permuted displacement method

1: draw $r_1 \sim \mathcal{U}(0, 1)$,
2: **for** $k \leftarrow 2, \ldots, \alpha - 1$ **do**
3: set $r_k \leftarrow \langle 2^{k-2} r_1 + 1/2 \rangle$,
4: **end for**
5: set $r_\alpha \leftarrow 1 - \langle 2^{\alpha-2} r_1 \rangle$,
6: pick a random $\sigma \in \mathsf{S}_\alpha$,
7: **for** $k \leftarrow 1, \ldots, \alpha$ **do**
8: set $U_k \triangleq r_{\sigma(k)}$,
9: **end for**

(7) in Algorithm 3 can be carried out by producing, using Algorithm 4, PNA uniforms $\{U_k\}_{k=1}^{\alpha}$ and letting, for $k \in \{1, \ldots, \alpha\}$ and $j \in \{1, \ldots, M_N\}$,

$$f(\tilde{\xi}_{N,\alpha(j-1)+k}) = F^{\leftarrow}_{\xi_{N,j}}[f](U_k),$$

where $F_\xi[f](x) \triangleq \mathcal{L}(\xi, \{f(\tilde{\xi}) \leq x\}) / \mathcal{L}(\xi, \tilde{\Xi})$, with $x \in \mathbb{R}$ and $\xi \in \Xi$, denotes the conditional distribution function of the $f(\tilde{\xi}_{N,\alpha(j-1)+k})$s given $\xi_{N,j} = \xi$. Since each function $F^{\leftarrow}_\xi[f]$ is monotone, it follows that $\{f(\tilde{\xi}_{N,\alpha(j-1)+k})\}_{k=1}^{\alpha}$ are conditionally EA. In the case $\tilde{\Xi} \subseteq \mathbb{R}$, a less *target function-specific* formulation of Algorithm 3 is possible: generating, using the permuted displacement method, PNA uniforms $\{U_k\}_{k=1}^{\alpha}$ and letting $\tilde{\xi}_{N,\alpha(j-1)+k} = F^{\leftarrow}_{\xi_{N,j}}[\mathrm{id}_{\tilde{\Xi}}](U_k)$ for all $k \in \{1, \ldots, \alpha\}$, with $\mathrm{id}_{\tilde{\Xi}}$ denoting the identity function on $\tilde{\Xi}$, yields conditionally EA variates $\{f(\tilde{\xi}_{N,\alpha(j-1)+k})\}_{k=1}^{\alpha}$ for the *whole class* of *monotone* functions f; this will be illustrated in Sect. 4. The case where $\tilde{\Xi} \subseteq \mathbb{R}^d$ and the target functions depend only on a *single* component of $\tilde{\xi}$ can be treated analogously. Of course, the method described above is applicable only when $F_\xi[f]$ is easy to invert; this is however not always the case and in Sect. 4 we present some alternative techniques for introducing negative correlation between the offspring particles.

3 Theoretical results

In this section, we justify theoretically Algorithm 3 using results on triangular arrays obtained by Douc and Moulines (2008). The arguments rely on results describing the weak convergence of Algorithm 1 and Algorithm 2 in a rather general setting.

3.1 Some notation and definitions

From now on the quality of a weighted sample will be described in terms of the following asymptotic properties, adopted from Douc and Moulines (2008), where a set C of real-valued functions on Ξ is said to be *proper* if the following conditions hold: (i) C is a linear space; (ii) if $g \in \mathsf{C}$ and f is measurable with $|f| \leq |g|$, then $|f| \in \mathsf{C}$; (iii) for all $c \in \mathbb{R}$, the constant function $f \equiv c$ belongs to C.

Definition 2 (*consistency*) A weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ on Ξ is said to be *consistent* for the probability measure μ and the proper set C if, for any $f \in \mathsf{C}$, as $N \to \infty$,

$$\Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f(\xi_{N,i}) \xrightarrow{\mathbb{P}} \mu(f),$$
$$\Omega_N^{-1} \max_{1 \le i \le M_N} \omega_{N,i} \xrightarrow{\mathbb{P}} 0.$$

Definition 3 (*asymptotic normality*) A weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ on Ξ is called *asymptotically normal* (AN) for $(\mu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{a_N\}_{N=1}^{\infty})$ if A and W are proper and, as $N \to \infty$,

$$a_N \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} [f(\xi_{N,i}) - \mu(f)] \xrightarrow{\mathcal{D}} \mathsf{N}[0, \sigma^2(f)] \quad \text{for any } f \in \mathsf{A},$$
$$a_N^2 \Omega_N^{-1} \sum_{i=1}^{M_N} (\omega_{N,i})^2 f(\xi_{N,i}) \xrightarrow{\mathbb{P}} \gamma(f) \quad \text{for any } f \in \mathsf{W},$$
$$a_N \Omega_N^{-1} \max_{1 \le i \le M_N} \omega_{N,i} \xrightarrow{\mathbb{P}} 0.$$

(Here $\mathsf{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2).

We impose the following assumptions.

(A1) The initial sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ is consistent for (ν, C).

(A2) The initial sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ is AN for $(\nu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{a_N\}_{N=1}^{\infty})$.

Under (A1) and (A2), we define

$$\begin{aligned}
\tilde{\mathsf{C}} &\triangleq \{f \in \mathsf{L}^1(\tilde{\Xi}, \mu) : L(\cdot, |f|) \in \mathsf{C}\}, \\
\tilde{\mathsf{A}} &\triangleq \{f : L(\cdot, |f|) \in \mathsf{A}, \mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2) \in \mathsf{W}; k \in \{1, \dots, \alpha\}\}, \\
\tilde{\mathsf{W}} &\triangleq \{f : \mathcal{R}_{0,k}(\cdot, \Phi_k^2 |f|) \in \mathsf{W}; k \in \{1, \dots, \alpha\}\}.
\end{aligned} \tag{12}$$

Moreover, let, for $f \in \tilde{\mathsf{A}}$ and $\xi \in \Xi$, assuming that $m \le n$,

$$\begin{aligned}
\mathbb{M}_{m,n}(\xi, f) = \int_{\tilde{\Xi}} \cdots \int_{\tilde{\Xi}} &\mathcal{R}_{m,n}(\xi, \tilde{\xi}_{1:m}, \Phi_n(\xi, \cdot) f) \Phi_m(\xi, \tilde{\xi}_m) f(\tilde{\xi}_m) \\
&\otimes_{\ell=1}^{m} R_\ell(\xi, \mathrm{d}\tilde{\xi}_1 \times \cdots \times \mathrm{d}\tilde{\xi}_m).
\end{aligned} \tag{13}$$

Straightforwardly, by definition, $\mathbb{M}_{m,n}(\xi, f)$ is the conditional expectation of $\Phi_m(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+m}) \Phi_n(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+n}) f(\tilde{\xi}_{N,\alpha(j-1)+m}) f(\tilde{\xi}_{N,\alpha(j-1)+n})$ given $\xi_{N,j} = \xi$, and we introduce the conditional covariances

$$\begin{aligned}\mathbb{C}_{m,n}(\xi, f) &\triangleq \operatorname{Cov}\Big[\Phi_m(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+m}) f(\tilde{\xi}_{N,\alpha(j-1)+m}),\\ &\quad \times \Phi_n(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+n}) f(\tilde{\xi}_{N,\alpha(j-1)+n}) \mid \xi_{N,j} = \xi\Big]\\ &= \mathbb{M}_{m,n}(\xi, f) - L^2(\xi, f), \quad (f, \xi) \in \tilde{\mathsf{A}} \times \Xi. \end{aligned} \tag{14}$$

3.2 Theoretical properties of Algorithm 1 and Algorithm 2

Under the assumptions above, we have the following convergence results, whose proofs are found in Appendix 1.

Theorem 1 *Assume* (A1) *and suppose that* $L(\cdot, \tilde{\Xi}) \in \mathsf{C}$. *Then the set* $\tilde{\mathsf{C}}$ *defined in* (12) *is proper and the weighted sample* $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha M_N}$ *produced by Algorithm 1 is consistent for* $(\mu, \tilde{\mathsf{C}})$.

Theorem 2 *Let the assumptions of Theorem 1 hold. In addition, assume* (A2) *and suppose that all functions* $\mathcal{R}_{0,k}(\cdot, \Phi_k^2)$, $k \in \{1, \ldots, \alpha\}$, *belong to* W. *Moreover, assume that* $L(\cdot, \tilde{\Xi})$ *belongs to* A. *Then the sets* $\tilde{\mathsf{A}}$ *and* $\tilde{\mathsf{W}}$ *defined in* (12) *are proper and the weighted sample* $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha M_N}$ *produced by Algorithm 1 is AN for* $(\mu, \tilde{\mathsf{A}}, \tilde{\mathsf{W}}, \tilde{\sigma}, \tilde{\gamma}, \{a_N\}_{N=1}^{\infty})$, *where, for* $f \in \tilde{\mathsf{A}}$,

$$\tilde{\sigma}^2(f) \triangleq \sigma^2\{L[f-\mu(f)]\}/[\nu L(\tilde{\Xi})]^2 + \sum_{(m,n)\in\{1,\ldots,\alpha\}^2} \gamma \mathbb{C}_{m,n}[f-\mu(f)]/[\alpha\nu L(\tilde{\Xi})]^2, \tag{15}$$

and, for $f \in \tilde{\mathsf{W}}$,

$$\tilde{\gamma}(f) \triangleq \sum_{k=1}^{\alpha} \gamma \mathcal{R}_{0,k}\left(\Phi_k^2 f\right)/[\alpha\nu L(\tilde{\Xi})]^2. \tag{16}$$

Remark 1 In the case where $R_k(\xi, \tilde{\xi}_{i:k-1}, \cdot) = R(\xi, \cdot)$ and $\Phi_k = \Phi = \mathrm{d}L/\mathrm{d}R$, that is, the particles within a block are mutated independently of each other, we have that $\mathbb{C}_{m,n} = 0$ for all $m \neq n$. This yields an asymptotic variance (15) of form

$$\begin{aligned}\tilde{\sigma}^2(f) &= \sigma^2\{L[f-\mu(f)]\}/[\nu L(\tilde{\Xi})]^2 + \sum_{m=1}^{\alpha} \gamma \mathbb{C}_{m,m}[f-\mu(f)]/[\alpha\nu L(\tilde{\Xi})]^2\\ &= \sigma^2\{L[f-\mu(f)]\}/[\nu L(\tilde{\Xi})]^2 + \alpha^{-1}\{\gamma R(\Phi^2[f-\mu(f)]^2)\\ &\quad - \gamma L^2[f-\mu(f)]\}/[\nu L(\tilde{\Xi})]^2, \end{aligned} \tag{17}$$

which is exactly the expression obtained by Douc and Moulines (2008, Theorem 2).

We move on to the convergence of Algorithm 2. Throughout the rest of this paper assume, entirely in line with Algorithm 3, that the adjustment multiplier weights satisfy the following assumption.

(A3) There exists a function $\Psi : \Xi \to \mathbb{R}^+$ such that $\psi_{N,i} = \Psi(\xi_{N,i})$ and $\Psi \in \mathsf{C} \cap \mathsf{L}^1(\Xi, \nu)$.

Define

$$\begin{aligned}
\bar{\mathsf{C}} &\triangleq \{f \in \mathsf{L}^1(\mu, \tilde{\Xi}) : L(\cdot, |f|) \in \mathsf{C} \cap \mathsf{L}^1(\nu, \tilde{\Xi})\},\\
\bar{\mathsf{A}} &\triangleq \big\{\Psi^{-1} L^2(\cdot, |f|) \in \mathsf{C} \cap \mathsf{L}^1(\nu, \Xi), L(\cdot, |f|) \in \mathsf{A}, L^2(\cdot, |f|) \in \mathsf{W},\\
&\qquad \Psi^{-1} \mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2) \in \mathsf{C} \cap \mathsf{L}^1(\nu, \Xi); k \in \{1, \ldots, \alpha\}\big\},\\
\bar{\mathsf{W}} &\triangleq \big\{\Psi^{-1} \mathcal{R}_{0,k}(\cdot, \Phi_k^2 |f|) \in \mathsf{C} \cap \mathsf{L}^1(\nu, \Xi); k \in \{1, \ldots, \alpha\}\big\}.
\end{aligned} \tag{18}$$

Now, by combining Theorem 2 with results obtained by Douc et al. (2008) we establish the convergence of Algorithm 2. This is the contents of the following corollaries whose proofs, which are obtained along the lines of Douc et al. (2008, Theorem 3.1), are omitted for brevity.

Corollary 1 *Let the assumptions of Theorem 1 hold and assume* (A3). *Then the set* $\bar{\mathsf{C}}$ *defined in* (18) *is proper and the weighted sample* $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha \tilde{M}_N}$ *obtained in Algorithm 2 is consistent for* $(\mu, \bar{\mathsf{C}})$.

Corollary 2 *Let the assumptions of Theorem 1 hold and assume* (A2) *with* $a_N^2 / M_N \to \beta$, $\beta \in [0, \infty)$. *In addition, suppose that* $\Psi \in \mathsf{A}$, $\Psi^2 \in \mathsf{W}$ *and that all functions* $\Psi^{-1} \mathcal{R}_{0,k}(\cdot, \Phi_k^2)$, $k \in \{1, \ldots, \alpha\}$, *belong to* $\mathsf{C} \cap \mathsf{L}^1(\nu, \tilde{\Xi})$. *Moreover, assume that* $\Psi^{-1} L^2(\cdot, \tilde{\Xi}) \in \mathsf{C} \cap \mathsf{L}^1(\nu, \tilde{\Xi})$, $L(\cdot, \tilde{\Xi}) \in \mathsf{A}$, *and* $L^2(\cdot, \tilde{\Xi}) \in \mathsf{W}$. *Then the sets* $\bar{\mathsf{A}}$ *and* $\bar{\mathsf{W}}$ *defined in* (18) *are proper and the weighted sample* $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha \tilde{M}_N}$ *obtained by Algorithm 2 with* $\tilde{M}_N / M_N \to \ell$, $\ell \in [0, \infty]$, *is AN for* $(\mu, \bar{\mathsf{A}}, \bar{\mathsf{W}}, \bar{\sigma}, \bar{\gamma}, \{a_N\}_{N=1}^{\infty})$, *where, for* $f \in \bar{\mathsf{A}}$,

$$\begin{aligned}
&\bar{\sigma}^2[\Psi](f) \triangleq \sigma^2\{L[\cdot, f - \mu(f)]\}/[\nu L(\tilde{\Xi})]^2\\
&\quad + \beta \ell^{-1} \nu(\Psi) \sum_{(m,n) \in \{1, \ldots, \alpha\}^2} \nu(\Psi \mathbb{M}_{m,n}\{\cdot, \Psi^{-1}[f - \mu(f)]\})/[\alpha \nu L(\tilde{\Xi})]^2
\end{aligned} \tag{19}$$

and, for $f \in \bar{\mathsf{W}}$,

$$\bar{\gamma}[\Psi](f) \triangleq \beta \ell^{-1} \nu(\Psi) \sum_{k=1}^{\alpha} \nu[\Psi^{-1} \mathcal{R}_{0,k}(\cdot, \Phi_k^2 f)]/[\alpha \nu L(\tilde{\Xi})]^2. \tag{20}$$

Remark 2 Note that as Algorithm 3 is only a special parametrisation of Algorithm 2, Corollary 1 and Corollary 2 imply also the consistency and AN of Algorithm 3, respectively.

Remark 3 The resampling operation in Step (5) in Algorithm 2 can, of course, be based on resampling techniques different from multinomial resampling, such as *residual resampling* or *Bernoulli branching*. The convergence results stated in Corollary 1 and Corollary 2 as well as the methodology developed above can be extended straightforwardly to these selection schemes, since their asymptotic behaviour is well investigated (see Chopin 2004; Douc and Moulines 2008).

3.3 Theoretical justification of Algorithm 3

To justify Algorithm 3, we examine the asymptotic variance (19), and check that successful antithetic coupling of the offspring particles of each block yields, in cases where the particle algorithm is close to fully adapted, indeed a reduction of asymptotic variance.

Antithetic mutation vs. independent mutation

We first study how the covariance structure within each block influences the asymptotic variance. Since the first term of (19) is not at all effected by the way the particles are mutated, we may focus entirely on the second term and write, using (14),

$$\begin{aligned}
&\beta\ell^{-1}\nu(\Psi) \sum_{(m,n)\in\{1,\ldots,\alpha\}^2} \nu(\Psi\mathbb{M}_{m,n}\{\cdot, \Psi^{-1}[f-\mu(f)]\})/[\alpha\nu L(\tilde{\Xi})]^2 \\
&\quad = \beta\ell^{-1}\nu(\Psi)\nu(\Psi L^2\{\cdot, \Psi^{-1}[f-\mu(f)]\})/[\nu L(\tilde{\Xi})]^2 \\
&\qquad + \beta\ell^{-1}\nu(\Psi) \sum_{(m,n)\in\{1,\ldots,\alpha\}^2} \nu(\Psi\mathbb{C}_{m,n}\{\cdot, \Psi^{-1}[f-\mu(f)]\})/[\alpha\nu L(\tilde{\Xi})]^2, \qquad (21)
\end{aligned}$$

where the first term on the RHS is again independent of the correlation structure of the mutation step. Compared to the case where the particles within each block are mutated conditionally independently (which implies that $\mathbb{C}_{m,n} = 0$ for all $m \neq n$), the variance will be smaller if the covariances $\mathbb{C}_{m,n}\{\cdot, \Psi^{-1}[f-\mu(f)]\}$ are negative for all $m \neq n$; however, when the algorithm is close to fully adapted, the product $\Psi^{-1}\Phi$ is close to unity, which implies that $\mathbb{C}_{m,n}\{\xi, \Psi^{-1}[f-\mu(f)]\}$ is close to the conditional covariance of $f(\tilde{\xi}_{N,\alpha(j-1)+m})$ and $f(\tilde{\xi}_{N,\alpha(j-1)+n})$ given $\xi_{N,j} = \xi$. Consequently, in the close to fully adapted case, we may expect that introducing, as in Algorithm 3, negative dependence between the $\{f(\tilde{\xi}_{N,\alpha(j-1)+k})\}_{k=1}^{\alpha}$ conditionally on $\xi_{N,j}$ leads to a significant decrease of variance compared to when the particles of each block are mutated independently.

Algorithm 3 vs. the standard APF

More interesting and relevant is to relate the performance of the antithetic SMC scheme in Algorithm 3 (with $\alpha > 1$) to that of the standard APF (for which $\alpha = 1$). More specifically, we wish to compare the following two updating procedures:

$$(1)\ \{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N} \xrightarrow{\text{sel.}} \{(\xi_{N,I_{N,i}}, 1)\}_{i=1}^{M_N} \xrightarrow{\text{mut. }(\alpha=1)} \{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{M_N},$$

$$(2)\ \{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N} \xrightarrow{\text{sel.}} \{(\xi_{N,I_{N,i}}, 1)\}_{i=1}^{\lceil M_N/\alpha\rceil} \xrightarrow{\text{mut. }(\alpha>1)} \{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{M_N},$$

where Procedure (1) corresponds to a full update of a close to fully adapted standard APF and Procedure (2) corresponds to an update of Algorithm 3. Note that Procedure (2) is expected to be computationally *more efficient* than (1), as (2) resamples only a fraction $\tilde{M}_N = \lceil M_N/\alpha\rceil$ of the original particle sample of size M_N at the selection step (in order to output a particle sample of the same size M_N as the input sample). The asymptotic variance of the standard APF (1) is provided by Corollary 2 with $\alpha = 1$ and $\ell = 1$:

$$\sigma^2_{(1)}(f) = \bar{\sigma}^2[\Psi](f)\Big|_{\ell=\alpha=1} = \sigma^2\{L[\cdot, f - \mu(f)]\}/[\nu L(\tilde{\Xi})]^2 \\ + \beta\nu(\Psi)\nu(\Psi\mathbb{M}_{1,1}\{\cdot, \Psi^{-1}[f - \mu(f)]\})/[\nu L(\tilde{\Xi})]^2. \quad (22)$$

We repeat again that Algorithm 3 is, as emphasised in Remark 2, a special parametrisation of Algorithm 2; thus, the asymptotic variance of Procedure (2), which we denote by $\sigma^2_{(2)}$, is provided by Corollary 2 with $\ell = 1/\alpha$:

$$\sigma^2_{(2)}(f) = \bar{\sigma}^2[\Psi](f)\Big|_{\ell=1/\alpha} = \sigma^2\{L[\cdot, f - \mu(f)]\}/[\nu L(\tilde{\Xi})]^2 \\ + \beta\alpha\nu(\Psi) \sum_{(m,n)\in\{1,\ldots,\alpha\}^2} \nu(\Psi\mathbb{M}_{m,n}\{\cdot, \Psi^{-1}[f - \mu(f)]\})/[\alpha\nu L(\tilde{\Xi})]^2. \quad (23)$$

Using these expressions, we aim at establishing some criterion (depending on the model as well as the objective function f under consideration) guaranteeing that Procedure (2) yields indeed a more accurate (in terms of asymptotic variance) estimator than the standard APF (1). However, comparing directly $\sigma^2_{(1)}(f)$ and $\sigma^2_{(2)}(f)$ under the assumption that the inherent covariance structure of each block is *uniform* yields the following criterion.

Corollary 3 *Assume that $\mathbb{M}_{m,n} = \mathbb{M}^*$ for all $(m, n) \in \{1, \ldots, \alpha\}^2$ such that $m \neq n$. Then*

$$\sigma^2_{(2)}(f) \le \sigma^2_{(1)}(f) \\ \Leftrightarrow \\ \nu(\Psi\mathbb{C}^*\{\cdot, \Psi^{-1}[f - \mu(f)]\}) \le -\nu(\Psi L^2\{\cdot, \Psi^{-1}[f - \mu(f)]\}), \quad (24)$$

where $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$, defined in (22) and (23), respectively, are the asymptotic variances associated with the Procedures (1) and (2) above, respectively.

The message provided by the criterion (24) is clear: by reducing the number of selected particles to the benefit of an increased number of antithetically coupled mutated offspring, Algorithm 3 can improve over a standard APF in a close to fully adaptive setting only when the conditional correlation between the mutated particles of each block is, on average under the *Ψ-modulated measure* $\nu\langle\Psi\rangle : \mathsf{A} \mapsto \nu(\Psi\mathbb{1}_{\mathsf{A}})/\nu(\Psi)$ (which can be shown to be the target distribution of the selection operation; see Douc et al. 2008 for details), lower than the second moment of $L\{\cdot, \Psi^{-1}[f - \mu(f)]\}$ under $\nu\langle\Psi\rangle$ with negative sign.

Remark 4 From the criterion (24), it is evident that imposing a nonnegative correlation structure among the particles in each block (that is, letting $\mathbb{C}^* \ge 0$) will, not surprisingly, *increase* the asymptotic variance vis-à-vis the standard APF. Moreover, since the correlation tends to zero with α, we conclude that there is a critical block size above which (24) will not hold even if the offspring particles of a block have the EA property conditionally on their ancestor.

4 Application to state-space models

In state-space models, a time series $Y \triangleq \{Y_n\}_{n=0}^{\infty}$, taking values in some state space $(\mathsf{Y}, \mathcal{B}(\mathsf{Y}))$, is modelled as noisy *observation* of an unobservable (possibly time-inhomogenous) Markov chain $X \triangleq \{X_n\}_{n=0}^{\infty}$. The Markov chain, also referred to as the *state sequence*, is assumed to take values in some state space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$. In the examples discussed below, we will exclusively let $\mathsf{X} \equiv \mathbb{R}$. The observed values are assumed to be conditionally independent given the latent process X in such a way that the distribution of each observation Y_n depends on the corresponding state X_n only. For a model of this type, any inference concerning the hidden states has to be carried through on the basis of the observations only.

Denote by $\{Q_n\}_{n=0}^{\infty}$ and ν_0 the Markov transition kernel and initial distribution of the hidden chain, respectively. In addition, suppose that the conditional distribution of Y_n given X_n admits a density g_n on Y with respect to some reference measure η, that is,

$$\mathbb{P}(Y_n \in \mathsf{A} \mid X_n) = \int_{\mathsf{A}} g_n(X_n, y)\, \eta(\mathrm{d}y), \quad \mathsf{A} \in \mathcal{B}(\mathsf{Y}).$$

This gives us a the following complete description of a state-space model:

$$\begin{aligned} X_0 &\sim \nu_0, \\ X_{n+1} \mid X_n &\sim Q_n(X_n, \cdot), \\ Y_n \mid X_n &\sim g_n(X_n, \cdot). \end{aligned} \tag{25}$$

4.1 Optimal filtering

In the setting of (25), the optimal filtering problem consists in computing, recursively in time as new observations become available, the *filter* posterior distributions

$$\phi_n(\mathsf{A}) \triangleq \mathbb{P}(X_n \in \mathsf{A} \mid Y_{0:n}), \quad \mathsf{A} \in \mathcal{B}(\mathsf{X}),\ n \geq 0. \tag{26}$$

One may establish (see, e.g. Cappé et al. 2005, Proposition 3.2.5) the recursion

$$\begin{aligned} \phi_0(\mathsf{A}) &= \frac{\int_{\mathsf{A}} g_0(x, Y_0)\, \nu_0(\mathrm{d}x)}{\int_{\mathsf{X}} g_0(x, Y_0)\, \nu_0(\mathrm{d}x)}, \quad \mathsf{A} \in \mathcal{B}(\mathsf{X}), \\ \phi_{n+1}(\mathsf{A}) &= \frac{\int_{\mathsf{X}} \int_{\mathsf{A}} g_{n+1}(x', Y_{n+1})\, Q_n(x, \mathrm{d}x')\, \phi_n(\mathrm{d}x)}{\iint_{\mathsf{X}^2} g_{n+1}(x', Y_{n+1})\, Q_n(x, \mathrm{d}x')\, \phi_n(\mathrm{d}x)}, \quad \mathsf{A} \in \mathcal{B}(\mathsf{X}), \end{aligned} \tag{27}$$

referred to as the *filtering recursion*. Since closed-form solutions to the filtering recursion are obtainable only in the case of a linear/Gaussian model or when the state space

X is finite, we follow Gordon et al. (1993) and apply the SMC methodology described in the previous; indeed, having defined, for $\mathsf{A} \in \mathcal{B}(\mathsf{X})$ and $x \in \mathsf{X}$, the unnormalised transition kernels

$$L_n(x, \mathsf{A}) = \int_{\mathsf{A}} g_{n+1}(x', Y_{n+1})\, Q_n(x, \mathrm{d}x'), \tag{28}$$

yielding the equivalent Feynman–Kac representation

$$\phi_{n+1}(\mathsf{A}) = \frac{\phi_n L_n(\mathsf{A})}{\phi_n L_n(\mathsf{X})}, \quad \mathsf{A} \in \mathcal{B}(\mathsf{X}),$$

of (38), we conclude that the optimal filtering problem can be perfectly cast into the framework of Sect. 2 with $\Xi = \tilde{\Xi} = \mathsf{X}$, $\nu = \phi_n$, $L = L_n$, and $\mu = \phi_{n+1}$.

Example 1 (*ARCH model*) As a first example, we consider the classical Gaussian *autoregressive conditional heteroscedasticity* (ARCH) *model* observed in noise (Bollerslev et al. 1994) given by

$$\begin{aligned} X_{n+1} &= W_{n+1}\sqrt{\beta_0 + \beta_1 X_n^2}, \\ Y_n &= X_n + \sigma V_n, \end{aligned}$$

where $\{W_n\}_{n=1}^{\infty}$ and $\{V_n\}_{n=0}^{\infty}$ are mutually independent sequences of standard normally distributed variables such that W_n is independent of $\{(X_k, Y_k)\}_{k=0}^{n}$ and V_n is independent of $\{(X_k, Y_k)\}_{k=0}^{n-1}$ and X_n. In this case, the optimal kernel $L_n(x, \cdot)/L_n(x, \mathsf{X})$, $x \in \mathbb{R}$, which in the state-space model setting is the conditional distribution of the state X_{n+1} given $X_n = x$ and the observation Y_{n+1}, is Gaussian with mean $m_n(x)$ and variance $\hat{\sigma}_n^2(x)$, where

$$m_n(x) = \frac{\beta_0 + \beta_1 x^2}{\beta_0 + \beta_1 x^2 + \sigma^2} Y_{n+1}, \quad \hat{\sigma}_n^2(x) = \frac{\beta_0 + \beta_1 x^2}{\beta_0 + \beta_1 x^2 + \sigma^2} \sigma^2.$$

Thus, the optimal adjustment multiplier weight function $\Psi_n(x) = L_n(x, \mathsf{X})$ can be expressed in a closed form as

$$\Psi_n(x) = \mathsf{N}(Y_{n+1}; 0, \beta_0 + \beta_1 x^2 + \sigma^2), \tag{29}$$

where $\mathsf{N}(x; \mu, \sigma^2) \triangleq \exp[-(x-\mu)^2/(2\sigma^2)]/\sqrt{2\pi\sigma^2}$ denotes (with a slight abuse of notation) the univariate Gaussian density function, yielding exactly uniform importance weights $\tilde{\omega}_{N,i} \equiv 1$, $i \in \{1, \ldots, \alpha M_N\}$.

In this setting, we used SMC to estimate posterior filter means $\{\phi_n(\mathrm{id}_{\mathsf{X}})\}_{n=0}^{30}$, where id_{X} denotes the identity mapping $\mathrm{id}_{\mathsf{X}}(x) = x$ on X. Initially, to form an idea of the effect of the antithetic coupling we compared the auxiliary particle filter in Algorithm 2, using $\alpha \in \{2, 3\}$ conditionally independent offspring of each particle $\xi_{N,i}$, $i \in \{1, \ldots, M_N\}$, in the mutation step, to the filter in Algorithm 3 using equally many antithetically coupled offspring. In the case $\alpha = 2$, we used the standard coupling

$$
\begin{aligned}
\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1} &= m_n(\xi^{(n)}_{N,i}) + \hat{\sigma}_n(\xi^{(n)}_{N,i})\epsilon^{(n)}_i, \\
\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+2} &= 2m_n(\xi^{(n)}_{N,i}) - \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1},
\end{aligned}
\tag{30}
$$

where $\{\epsilon^{(n)}_i\}_{i=1}^{M_N}$ is a sequence of mutually independent standard Gaussian random variables being independent of everything else. This coupling yields largest possible negative correlation (that is, is EA) conditionally on $\xi^{(n)}_{N,i}$, i.e. $\mathrm{Corr}(\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1}, \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+2} \mid \xi^{(n)}_{N,i}) = -1$, and in the kernel language of Sect. 2 it holds that $R_1(\xi, \mathsf{A}) = \int_{\mathsf{A}} \mathsf{N}(\tilde{\xi}; m_n(\xi), \hat{\sigma}^2_n(\xi))\, \mathrm{d}\tilde{\xi}$ and $R_2(\xi, \tilde{\xi}_1, \mathsf{A}) = \delta_{2m_n(\xi)-\tilde{\xi}_1}(\mathsf{A})$ for any Borel set $\mathsf{A} \subseteq \mathbb{R}$. A similar coupling was used in the case where $\alpha = 3$; here we set

$$
\begin{aligned}
\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1} &= m_n(\xi^{(n)}_{N,i}) + \hat{\sigma}_n(\xi^{(n)}_{N,i})\epsilon^{(n)}_{i,1}, \\
\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+2} &= \frac{1}{2}\left(3m_n(\xi^{(n)}_{N,i}) - \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1} + \sqrt{3}\hat{\sigma}_n(\xi^{(n)}_{N,i})\epsilon^{(n)}_{i,2}\right), \\
\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+3} &= 3m_n(\xi^{(n)}_{N,i}) - \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1} - \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+2},
\end{aligned}
\tag{31}
$$

where the independent sequences $\{\epsilon^{(n)}_{i,1}\}_{i=1}^{M_N}$ and $\{\epsilon^{(n)}_{i,2}\}_{i=1}^{M_N}$ are as above. The coupling (31) yields $\mathrm{Corr}(\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+m}, \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+m'} \mid \xi^{(n)}_{N,i}) = -1/2$, for $(m, m') \in \{1, 2, 3\}$ and $m \neq m'$.

The comparison was done for two different data sets obtained by simulation of ARCH models parametrised by $(\beta_0, \beta_1, \sigma) = (0.9, 0.6, 1)$ and $(\beta_0, \beta_1, \sigma) = (0.9, 0.6, 10)$, corresponding to informative and non-informative observations, respectively. The mean squared errors (MSEs) for 400 runs of each filter with $M_N = 6000/\alpha$ are, for the different values of α, displayed in Fig. 1a (the informative case) and Fig. 1b (the non-informative case). The MSEs are based on reference posterior filter mean values obtained by means of the standard APF (for which $\alpha = \ell = 1$) using as many as 500,000 particles. From both figures, it is evident that letting the particles of a block be antithetically coupled instead of conditionally independent decreases significantly the variance. Moreover, the improvement is especially noticeable in the informative case.

More relevant is to compare the performance of Algorithm 3, again with $\alpha \in \{2, 3\}$ and $M_N = 6000/\alpha$, to that of the standard fully adapted APF using 6000 particles without any block structure. In this setting, both antithetic filters are clearly more computationally efficient since, first, only a half and a third of the particles are selected at each resampling operation, and, second, a half and a third of the random moves at each mutation step are replaced by simple assignments (matrix manipulations) in the two cases $\alpha = 2$ and $\alpha = 3$, respectively. The outcome is displayed in Fig. 2 from which it is clear that performances of the antithetic filters are, despite being less costly, superior, especially in the case of informative observations (see Fig. 2a); indeed, the improvement is over 20 decibel at some time steps. Moreover, it is evident that the

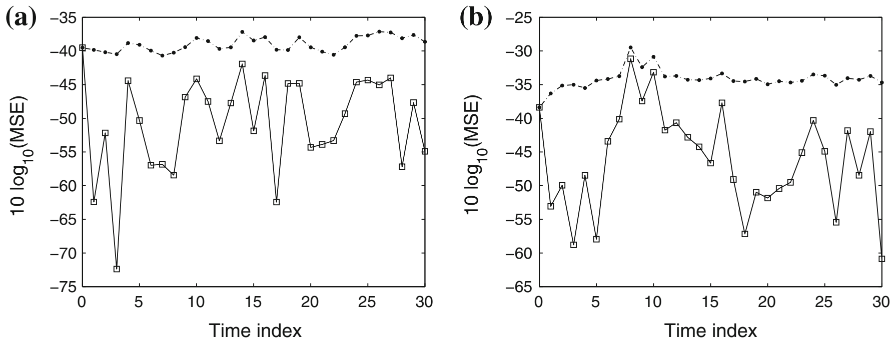

Fig. 1 Plot of MSEs (in decibel) of filters being implementations of Algorithm 3 with $\alpha = 2$ antithetically coupled (*open square*) and conditionally independent (*open circle*) offspring for the ARCH model with informative (**a**) and non-informative (**b**) observations. The MSE values are based on 400 runs of each algorithm with $M_N = 3000$

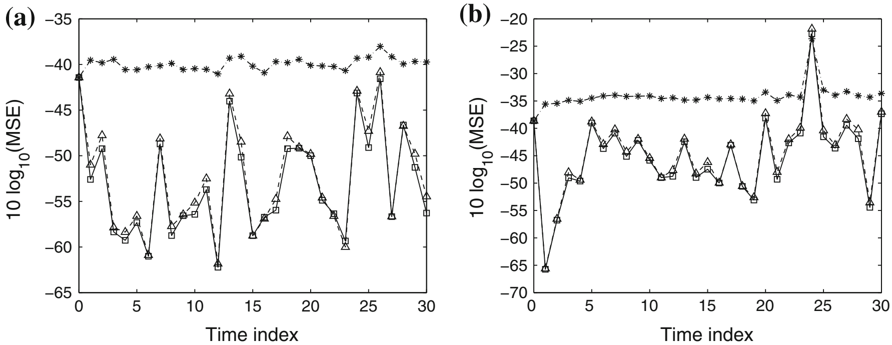

Fig. 2 Plot of MSEs (in decibel) of the standard optimal APF (*asterisk*) with 6000 particles and antithetic filters with $\alpha = 2$ (*open square*) and $\alpha = 3$ (*open triangle*) for the ARCH model with informative (**a**) and non-informative (**b**) observations. $\alpha M_N = 6000$ for both antithetic filters and the MSE values are based on 400 runs of each algorithm

computational gain of using $\alpha = 3$ instead of $\alpha = 2$ offspring in each block is at the expense of a slight decrease of precision.

Remark 5 The noisily observed ARCH model discussed above belongs to the larger class of nonlinear/Gaussian state-space models of form

$$\begin{aligned} X_{n+1} &= a_n(X_n) + b_n(X_n) W_{n+1}, \\ Y_n &= b X_n + s V_n, \end{aligned}$$

where $\{W_n\}_{n=1}^{\infty}$ and $\{V_n\}_{n=0}^{\infty}$ are as in Example 1 and $\{a_n\}_{n \geq 0}$, $\{b_n\}_{n \geq 0}$ and b, s are sequences of matrix-valued functions and matrices, respectively, of appropriate dimensions. As mentioned above, the optimal kernel and adjustment multiplier weights can be evaluated on closed form for models of this sort, and we refer again to Cappé et al. (2005, Section 7.2.2.2) for details.

Example 2 (*Growth model*) The univariate *growth model* given by, for $n \geq 0$,

$$
\begin{aligned}
X_{n+1} &= a_n(X_n) + \sigma_w W_{n+1}, \\
Y_n &= b X_n^2 + \sigma_v V_n,
\end{aligned}
\tag{32}
$$

where

$$
a_n(x) = \alpha_0 x + \alpha_1 \frac{x}{1+x^2} + \alpha_2 \cos(1.2n), \quad x \in \mathbb{R},
$$

and the sequences $\{W_n\}_{n=1}^{\infty}$ and $\{V_n\}_{n=0}^{\infty}$ are as in the previous example, was discussed by Kitagawa (1987) and has served as a benchmark for state-space filtering techniques during the last decades. We will follow the lines of Cappé et al. (2005) and consider the parameter vector $(\alpha_0, \alpha_1, \alpha_2, b, \sigma_v^2) = (0.5, 25, 8, 0.05, 1)$ and $\sigma_w^2 \in \{1, 10\}$, the values of the latter parameter corresponding to non-informative and informative observations, respectively. The initial state is set deterministically to $X_0 = 0.1$. For a given observation Y_n in $\mathbb{R}$, the local likelihood for the state at time n is given by the function

$$
\mathbb{R} \ni x \mapsto g(x, Y_n) = \mathsf{N}(Y_n; bx^2, \sigma_v^2), \tag{33}
$$

which is symmetric around zero for any observation Y_n. Interestingly, functions (33) associated with negative observations $Y_n \leq 0$ are *unimodal*, while those associated with positive observations $Y_n > 0$ are *bimodal* with modes located at $\pm\sqrt{Y_n/b}$. This bimodality is challenging from a filtering point of view and puts heavy demands on the applied SMC method.

Unlike the ARCH model in the previous section, direct simulation from the optimal kernel is infeasible in this case since the measurement Eq. (32) is nonlinear in the state. Thus, in order to mimic efficiently the optimal kernel and adjustment multiplier weights we take a novel approach and approximate the local likelihood (33) by a mixture

$$
\mathcal{G}(x, Y_n) \triangleq \mathsf{N}(x; \mu_1(Y_n), \varsigma^2(Y_n))/2 + \mathsf{N}(x; \mu_2(Y_n), \varsigma^2(Y_n))/2, \quad x \in \mathbb{R},
$$

of two Gaussian densities, where

$$
(\mu_1(Y_n), \mu_2(Y_n), \varsigma^2(Y_n)) \triangleq \begin{cases} (0, 0, -\sigma_v^2/(2bY_n)) & \text{for} \quad Y_n \leq 0, \\ (-\sqrt{Y_n/b}, \sqrt{Y_n/b}, \sigma_v^2/(4bY_n)) & \text{for} \quad Y_n > 0. \end{cases}
$$

Consequently, we let the means and standard deviations of the two strata be the locations (which coincide when $Y_n \leq 0$) and (common) inverted log curvature (with negative sign) of the modes of the local likelihood, respectively; more specifically, $\varsigma^2(Y_n) = -1/(\mathrm{d}^2 \log g(x, Y_n)/\mathrm{d}x^2)|_{x=\mu_1(Y_n)}$. From now on, we omit for brevity the dependence on the observation from the notation of the quantities above and write $(\mu_1, \mu_2, \varsigma^2)$ instead of $(\mu_1(Y_n), \mu_2(Y_n), \varsigma^2(Y_n))$. Plugging the approximation $\mathcal{G}$ into the expression (28) of the unnormalised optimal kernel yields straightforwardly the mixture

$$\begin{aligned}\mathcal{L}_n(x, \mathsf{A}) &\triangleq \int_{\mathsf{A}} \mathcal{G}(x', Y_{n+1})\, Q_n(x, \mathrm{d}x') \\ &= \beta_n^{(1)}(x) G_n^{(1)}(x, \mathsf{A}) + \beta_n^{(2)}(x) G_n^{(2)}(x, \mathsf{A}), \quad x \in \mathsf{X}, \quad \mathsf{A} \in \mathcal{B}(\mathsf{X}),\end{aligned}$$

where each Gaussian stratum

$$G_n^{(d)}(x, \mathsf{A}) \triangleq \int_{\mathsf{A}} \mathsf{N}\left(x'; \tau_n^{(d)}(x), \eta_n^2\right) \mathrm{d}x', \quad d \in \{1, 2\},$$

with means and variance (recall that μ_d, $d \in \{1, 2\}$, and ς^2 depend on Y_{n+1})

$$\begin{aligned}\tau_n^{(d)}(x) &\triangleq \frac{\sigma_w^2 \mu_d + \varsigma^2 a_n(x)}{\sigma_w^2 + \varsigma^2}, \\ \eta_n^2 &\triangleq \frac{\sigma_w^2 \varsigma^2}{\sigma_w^2 + \varsigma^2},\end{aligned}$$

is weighted by

$$\beta_n^{(d)}(x) \triangleq \mathsf{N}(\mu_d; a_n(x), \sigma_w^2 + \varsigma^2), \quad d \in \{1, 2\}.$$

By normalising, we obtain the approximation

$$\mathcal{L}_n(x, \mathsf{A}) / \mathcal{L}_n(x, \mathsf{X}) = \bar{\beta}_n(x) G_n^{(1)}(x, \mathsf{A}) + (1 - \bar{\beta}_n(x)) G_n^{(2)}(x, \mathsf{A}), \quad x \in \mathsf{X}, \ \mathsf{A} \in \mathcal{B}(\mathsf{X}), \tag{34}$$

of the optimal kernel, where we have defined the normalised weight

$$\bar{\beta}_n(x) \triangleq \frac{\beta_n^{(1)}(x)}{\beta_n^{(1)}(x) + \beta_n^{(2)}(x)}, \quad x \in \mathsf{X}.$$

Moreover, in this setting, the approximate optimal adjustment multiplier weights are given by

$$\Psi_n(x) = \mathcal{L}_n(x, \mathsf{X}) = \beta_n^{(1)}(x) + \beta_n^{(2)}(x), \quad x \in \mathsf{X}.$$

Using (34) as proposal, the experiment of the previous example (in which we estimated filter posterior means $\{\phi_n(\mathrm{id}_{\mathsf{X}})\}_{n=0}^{30}$) was repeated with focus set on the case $\alpha = 2$. To impose a conditionally negative correlation structure, we let each pair of offspring particles evolve according to

$$\begin{aligned}\tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)} &= \tau_n^{(1)}(\xi_{N,i}^{(n)}) \mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi_{N,i}^{(n)})\}} + \tau_n^{(2)}(\xi_{N,i}^{(n)}) \mathbb{1}_{\{U_i^{(n)} \geq \bar{\beta}_n(\xi_{N,i}^{(n)})\}} + \eta_n \epsilon_i^{(n)}, \\ \tilde{\xi}_{N,\alpha(i-1)+2}^{(n+1)} &= \tau_n^{(1)}(\xi_{N,i}^{(n)}) \mathbb{1}_{\{1-U_i^{(n)} < \bar{\beta}_n(\xi_{N,i}^{(n)})\}} + \tau_n^{(2)}(\xi_{N,i}^{(n)}) \mathbb{1}_{\{1-U_i^{(n)} \geq \bar{\beta}_n(\xi_{N,i}^{(n)})\}} - \eta_n \epsilon_i^{(n)},\end{aligned} \tag{35}$$

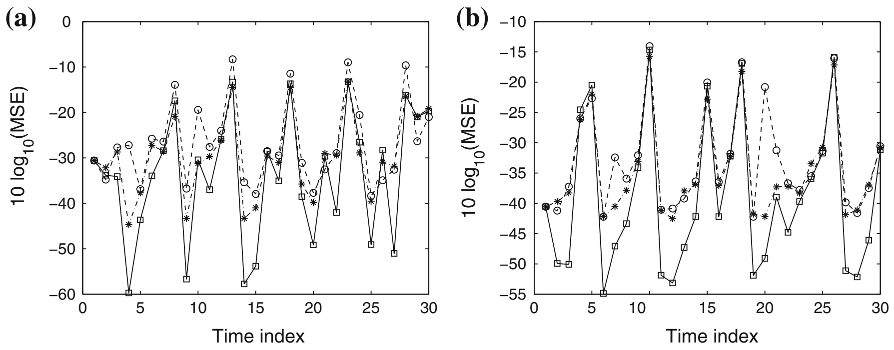

Fig. 3 Plot of MSEs (in decibel) of the plain bootstrap filter (*open circle*) using 5000 particles, the standard optimal APF (*asterisk*) using 5000 particles, and the antithetic filter (*open square*) with $\alpha = 2$ and $\alpha M_N = 5000$ for the growth model with informative (**a**) and non-informative (**b**) observations. The MSE values are based on 400 runs of each algorithm

where $\{U_i^{(n)}\}_{i=1}^{M_N}$ and $\{\epsilon_i^{(n)}\}_{i=1}^{M_N}$ are independent sequences of mutually independent uniformly distributed (on $[0, 1]$) and standard Gaussian random variables, respectively, such that each pair $(U_i^{(n)}, \epsilon_i^{(n)})$ is independent of all other random variables. It is established easily that each of the offspring particles $\tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)}$ and $\tilde{\xi}_{N,\alpha(i-1)+2}^{(n+1)}$ of the copuling (35) is distributed marginally according to the approximate optimal kernel (34). In addition, one can show that (see Sect. 1 for details) the correlation between the offspring of a block is given by, for $\xi \in \mathsf{X}$,

$$\begin{aligned}&\mathrm{Corr}\left[\tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)}, \tilde{\xi}_{N,\alpha(i-1)+2}^{(n+1)} \mid \xi_{N,i}^{(n)} = \xi\right] \\ &\quad = -\frac{(\tau_n^{(1)}(\xi) - \tau_n^{(2)}(\xi))^2 \left[\bar{\beta}_n^2(\xi)\mathbb{1}\{\bar{\beta}_n(\xi) \le 1/2\} + (\bar{\beta}_n^2(\xi) - 1)^2 \mathbb{1}\{\bar{\beta}_n(\xi) > 1/2\}\right] + \eta_n^2}{(\tau_n^{(1)}(\xi) - \tau_n^{(2)}(\xi))^2 \bar{\beta}_n(\xi)(1 - \bar{\beta}_n(\xi)) + \eta_n^2},\end{aligned} \tag{36}$$

which is *always* negative and simplifies to -1 in the unimodal case (as $\tau_n^{(1)}(\xi) = \tau_n^{(2)}(\xi)$ for all $\xi \in \mathsf{X}$ when $Y_{n+1} < 0$). Figure 3 displays MSE (in decibel) comparisons between the antithetic APF with $\alpha = 2$ and $\alpha M_N = 5000$, a close to fully adapted APF, based on the proposal kernel (34) and 5000 particles, and the plain bootstrap filter using 5000 particles. Like in the ARCH example, we let the filters approximate filter posterior means $\phi_n(\mathrm{id}_{\mathsf{X}})$ for observation records of length 30, and since the initial value is known deterministically the log MSE is null at time zero. The comparison was made for informative ($\sigma_w^2 = 10$, Fig. 3a) as well as non-informative ($\sigma_w^2 = 1$, Fig. 3b) observations and the MSEs, measured with respect to reference values obtained with the close to fully adapted APF using 500,000 particles, were based on 400 runs of each algorithm. Also for this demanding model the variance reduction introduced by the antithetic coupling is significant; indeed, despite being clearly less computationally costly (see the discussion in the previous example), the antithetic filter improves the MSE performances of the APF and the bootstrap filter by more than 10 decibels at several time points for both observation records. Moreover, from the figures, it is evident that proposing particles according to the approximate optimal kernel (34) instead of the prior kernel yields, as expected, generally more precise posterior filter mean estimates, since the APF outperforms the bootstrap particle filter at most time steps.

4.2 Approximation of the predictor distribution flow

In many applications, the objects of interest are the *predictor distributions*

$$\pi_n(\mathsf{A}) \triangleq \mathbb{P}(X_n \in \mathsf{A} \mid Y_{0:n-1}), \quad \mathsf{A} \in \mathcal{B}(\mathsf{X}), \quad n \geq 0, \tag{37}$$

rather than the filter distributions. This is, e.g. the case when approximating the *log-likelihood function*

$$\ell_n(Y_{0:n}) = \sum_{k=0}^{n} \log \pi_k(g_k)$$

(with the convention $\pi_0 = \nu_0$) of the observed data $Y_{0:n}$. As for the filter distributions, one may derive a recursion

$$\begin{aligned} \pi_0 &= \nu_0, \\ \pi_{n+1}(\mathsf{A}) &= \frac{\int_{\mathsf{X}} g_n(x, Y_n)\, Q_n(x, \mathsf{A})\, \pi_n(\mathrm{d}x)}{\int_{\mathsf{X}} g_n(x, Y_n)\, \pi_n(\mathrm{d}x)}, \quad \mathsf{A} \in \mathcal{B}(\mathsf{X}), \end{aligned} \tag{38}$$

for these measures, which can, just like in the filtering case, can be cast into our general framework (5) by defining

$$\tilde{L}_n(x, \mathsf{A}) = g_n(x, Y_n) Q_n(x, \mathsf{A}), \quad (x, \mathsf{A}) \in \mathsf{X} \times \mathcal{B}(\mathsf{X}),$$

and letting $\Xi = \tilde{\Xi} = \mathsf{X}$, $\nu = \pi_n$, $L = \tilde{L}_n$, and $\mu = \pi_{n+1}$. In this case, perfectly full adaptation is straightforward as $\tilde{L}_n(x, \mathsf{X}) = g_n(x, Y_n)$ and $\tilde{L}_n(x, \cdot)/\tilde{L}_n(x, \mathsf{X}) = Q_n(x, \cdot)$, $x \in \mathsf{X}$, and the antithetic strategy proposed by us can be implemented easily for many models of interest. In fact, when considering a recursion for the predictor distributions, the fully adapted algorithm in the setting of Sect. 2.2 corresponds to the standard *bootstrap particle filter*, which mutates the particles according to the prior dynamics of the state sequence and weighs the same using the local likelihood.

5 Conclusion

The present paper casts antithetic acceleration into the framework of SMC methods by introducing, in the mutation operation of the SMC algorithm, negative dependence between blocks of particles. In a scenario of full adaptation, i.e. when the so-called optimal kernel $L(\xi, \cdot)/L(\xi, \tilde{\Xi})$ and the optimal adjustment weight function $L(\xi, \tilde{\Xi})$ can, at least approximately and pointwise for all $\xi \in \Xi$, be sampled from and computed, respectively, we have shown theoretically that introducing sufficiently strong correlation between the particles of each block lowers indeed the asymptotic variance added at each step of the SMC algorithm. These theoretical results were confirmed by a simulation study within the framework of optimal filtering in state-space models, which reported improvements of up to an order of magnitude for some models under consideration.

The fact that our approach requires the optimal kernel and adjustment multiplier weights to be known on closed form or to be well approximated may be viewed as drawback. Nevertheless, such approximation is in fact possible for a wide range of models, e.g. the nonlinear/Gaussian models of the type described in Remark 5 or the important case of approximation of the predictor distribution flow in state-space models.

A natural future research project aims at accelerate further the algorithm by extending the antithetic coupling to the selection operation.

A Proofs

A.1 Proof of Theorem 1

The result follows straightforwardly from Slutsky's theorem and results obtained by Douc and Moulines (2008) in the case of independently mutated particles. Indeed, by Douc and Moulines (2008, Equation (36)) we have, for any $k \in \{1, \ldots, \alpha\}$, as $N \to \infty$,

$$\Omega_N^{-1} \sum_{j=1}^{M_N} \tilde{\omega}_{N,\alpha(j-1)+k} f(\tilde{\xi}_{N,\alpha(j-1)+k}) \stackrel{\mathbb{P}}{\longrightarrow} \nu L(f),$$

yielding immediately

$$\begin{aligned}
&(\alpha \Omega_N)^{-1} \sum_{i=1}^{\alpha M_N} \tilde{\omega}_{N,i} f(\tilde{\xi}_{N,i}) \\
&\quad = \alpha^{-1} \sum_{k=1}^{\alpha} \Omega_N^{-1} \sum_{j=1}^{M_N} \tilde{\omega}_{N,\alpha(j-1)+k} f(\tilde{\xi}_{N,\alpha(j-1)+k}) \stackrel{\mathbb{P}}{\longrightarrow} \nu L(f). \qquad (39)
\end{aligned}$$

By applying (39) for this limit for $f \equiv 1$ (recall that $L(\cdot, \tilde{\Xi}) \in \mathsf{C}$ by assumption, implying that the constant function belongs to $\tilde{\mathsf{C}}$) we obtain, using again Slutsky's theorem,

$$\tilde{\Omega}_N^{-1} \sum_{i=1}^{\alpha M_N} \tilde{\omega}_{N,i} f(\tilde{\xi}_{N,i}) \stackrel{\mathbb{P}}{\longrightarrow} \nu L(f)/\nu L(\tilde{\Xi}) = \mu(f).$$

To prove the second property in Definition 2, write

$$(\alpha\Omega_N)^{-1} \max_{1\le i\le \alpha M_N} \tilde{\omega}_{N,i} \le \alpha^{-1} \sum_{k=1}^{\alpha} \Omega_N^{-1} \max_{1\le j\le M_N} \tilde{\omega}_{N,\alpha(j-1)+k}; \tag{40}$$

however, by inspecting the proof of Douc and Moulines (2008, Theorem 1) we conclude that each term on the RHS of (40) tends to zero in probability, which in combination with (39) implies that

$$\tilde{\Omega}_N^{-1} \max_{1\le i\le \alpha M_N} \tilde{\omega}_{N,i} = (\alpha\Omega_N/\tilde{\Omega}_N)(\alpha\Omega_N)^{-1} \max_{1\le i\le \alpha M_N} \tilde{\omega}_{N,i} \xrightarrow{\mathbb{P}} 0.$$

This completes the proof. □

A.2 Proof of Theorem 2

Let $f \in \tilde{\mathsf{A}}$ and assume without loss of generality that $\mu(f) = 0$. Then write, following the lines of the proof of Douc and Moulines (2008, Theorem 2)

$$a_N \tilde{\Omega}_N^{-1} \sum_{i=1}^{\alpha M_N} \tilde{\omega}_{N,i} f(\tilde{\xi}_{N,i}) = \alpha \Omega_N \tilde{\Omega}_N^{-1}(A_N + B_N),$$

where

$$A_N \triangleq \sum_{j=1}^{M_N} \mathbb{E}\left[U_{N,j} \mid \mathcal{F}_{N,\alpha(j-1)}\right], \quad B_N \triangleq \sum_{j=1}^{M_N} \left\{U_{N,j} - \mathbb{E}\left[U_{N,j} \mid \mathcal{F}_{N,\alpha(j-1)}\right]\right\},$$

and $U_{N,j} \triangleq a_N(\alpha\Omega_N)^{-1} \sum_{k=1}^{\alpha} \tilde{\omega}_{N,\alpha(j-1)+k} f(\tilde{\xi}_{N,\alpha(j-1)+k})$. Since, by (39), $\tilde{\Omega}_N/(\alpha\Omega_N) \xrightarrow{\mathbb{P}} \nu L(\tilde{\Xi})$, as $N \to \infty$, it is enough to prove that

$$A_N + B_N \xrightarrow{\mathcal{D}} \mathsf{N}\{0, \sigma^2[L(\cdot, f)] + \eta^2(f)\}, \tag{41}$$

where

$$\eta^2(f) \triangleq \alpha^{-2} \sum_{(m,n)\in\{1,\ldots,\alpha\}^2} \gamma \mathbb{C}_{m,n}(f).$$

For A_N it holds, since the weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ is AN for $(\mu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{a_N\}_{N=1}^{\infty})$ by assumption and $L(\cdot, f) \in \mathsf{A}$, that

$$A_N = a_N(\alpha\Omega_N)^{-1} \sum_{j=1}^{M_N} \sum_{k=1}^{\alpha} \mathbb{E}\left[\tilde{\omega}_{N,\alpha(j-1)+k} f(\tilde{\xi}_{N,\alpha(j-1)+k}) \mid \mathcal{F}_{N,\alpha(j-1)}\right]$$

$$= a_N \Omega_N^{-1} \sum_{j=1}^{M_N} \omega_{N,j} L(\xi_{N,j}, f) \xrightarrow{\mathcal{D}} \mathsf{N}\{0, \sigma^2[L(\cdot, f)]\}.$$

We now consider B_N and establish that, for any $u \in \mathbb{R}$,

$$\mathbb{E}\left[\exp(\mathrm{i}u B_N) | \mathcal{F}_{N,0}\right] \xrightarrow{\mathbb{P}} \exp(-u^2\eta^2(f)/2), \tag{42}$$

from which the result of the theorem follows. The proof of (42) consists in showing that the two conditions of Theorem 13 in Douc and Moulines (2008) are satisfied for the triangular array $\{(U_{N,j}, \mathcal{F}_{N,\alpha j})\}_{j=1}^{M_N}$.

For establishing condition (i) of the theorem in question, write

$$\begin{aligned}
&\sum_{j=1}^{M_N} \mathbb{E}\left[U_{N,j}^2 \mid \mathcal{F}_{N,\alpha(j-1)}\right] \\
&\quad = a_N^2 (\alpha \Omega_N)^{-2} \sum_{j=1}^{M_N} \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \mathbb{E}\left[\tilde{\omega}_{N,\alpha(j-1)+k} f(\tilde{\xi}_{N,\alpha(j-1)+k})\right. \\
&\qquad \left. \times\, \tilde{\omega}_{N,\alpha(j-1)+m} f(\tilde{\xi}_{N,\alpha(j-1)+m}) \,\middle|\, \mathcal{F}_{N,\alpha(j-1)}\right] \\
&\quad = \alpha^{-2} \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} a_N^2 \Omega_N^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \mathbb{M}_{k,m}(\xi_{N,j}, f).
\end{aligned} \tag{43}$$

However, for all $(k, m) \in \{1, \ldots, \alpha\}^2$, $\mathbb{M}_{k,m}(\cdot, f) \le \mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2) + \mathcal{R}_{0,m}(\cdot, \Phi_m^2 f^2) \in \mathsf{W}$; since W is proper, this implies (under (A2)) the limit

$$\alpha^{-2} \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} a_N^2 \Omega_N^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \mathbb{M}_{k,m}(\xi_{N,j}, f) \xrightarrow{\mathbb{P}} \alpha^{-2} \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \gamma \mathbb{M}_{k,m}(f). \tag{44}$$

Now consider

$$\begin{aligned}
&\sum_{j=1}^{M_N} \mathbb{E}^2\left[U_{N,j} \mid \mathcal{F}_{N,\alpha(j-1)}\right] \\
&\quad = a_N^2 (\alpha \Omega_N)^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \mathbb{E}^2\left[\sum_{k=1}^{\alpha} \Phi_k(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+k}) f(\tilde{\xi}_{N,\alpha(j-1)+k}) \,\middle|\, \mathcal{F}_{N,\alpha(j-1)}\right] \\
&\quad = a_N^2 \Omega_N^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 L^2(\xi_{N,j}, f);
\end{aligned} \tag{45}$$

here, for any $k \in \{1, \ldots, \alpha\}$, $L^2(\cdot, f) = \mathcal{R}^2_{0,k}(\cdot, \Phi_k f) \leq \mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2) \in \mathsf{W}$, and reusing the asymptotic normality of $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ yields

$$a_N^2 \Omega_N^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 L^2(\xi_{N,j}, f) \xrightarrow{\mathbb{P}} \gamma L^2(f). \tag{46}$$

Finally, by combining Eqs. (43)–(46) we conclude that

$$\begin{aligned}&\sum_{j=1}^{M_N} \left\{ \mathbb{E}\left[U_{N,j}^2 \mid \mathcal{F}_{N,\alpha(j-1)} \right] - \mathbb{E}^2\left[U_{N,j} \mid \mathcal{F}_{N,\alpha(j-1)} \right] \right\} \\ &\quad \xrightarrow{\mathbb{P}} \alpha^{-2} \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \gamma \mathbb{M}_{k,m}(f) - \gamma L^2(f) = \eta^2(f),\end{aligned}$$

which establishes condition (i).

It remains to check condition (ii), that is, for any $\varepsilon > 0$,

$$C_N \triangleq \sum_{j=1}^{M_N} \mathbb{E}\left[U_{N,j} \mathbb{1}_{\{|U_{N,j}| \geq \varepsilon\}} \mid \mathcal{F}_{N,\alpha(j-1)} \right] \xrightarrow{\mathbb{P}} 0.$$

Thus, argue along the lines of the proof of Douc and Moulines (2008, Theorem 2) and write, for any $C > 0$,

$$\begin{aligned}C_N &\leq a_N^2 (\alpha \Omega_N)^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \mathbb{M}_{k,m}\left(\xi_{N,j}, f \mathbb{1}_{\{|\sum_{k=1}^{\alpha} \Phi_k f| \geq C\}} \right) \\ &\quad + \mathbb{1}_{\{a_N (\alpha\Omega_N)^{-1} \max_i \omega_{N,i} \geq \varepsilon C^{-1}\}} \sum_{j=1}^{M_N} \mathbb{E}\left[U_{N,j}^2 \middle| \mathcal{F}_{N,\alpha(j-1)} \right].\end{aligned} \tag{47}$$

Under (A2) the indicator function of the second term on the RHS of (47) tends to zero in probability and since, for all $(k, m) \in \{1, \ldots, \alpha\}^2$, $\mathbb{M}_{k,m}(\cdot, f \mathbb{1}_{\{|\sum_{k=1}^{\alpha} \Phi_k f| \geq C\}}) \leq \mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2) + \mathcal{R}_{0,m}(\cdot, \Phi_m^2 f^2) \in \mathsf{W}$ we obtain

$$\begin{aligned}&a_N^2 (\alpha\Omega_N)^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \mathbb{M}_{k,m}\left(\xi_{N,j}, f \mathbb{1}_{\{|\sum_{k=1}^{\alpha} \Phi_k f| \geq C\}} \right) \\ &\quad \xrightarrow{\mathbb{P}} \alpha^{-2} \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \gamma \mathbb{M}_{k,m}\left(f \mathbb{1}_{\{|\sum_{k=1}^{\alpha} \Phi_k f| \geq C\}} \right).\end{aligned} \tag{48}$$

By dominated convergence, the RHS of (48) can be made arbitrarily small by taking C sufficiently large. Therefore, also condition (ii) is satisfied, implying the convergence (42). This establishes (41).

We turn to the second property of Definition 3 and show that, for any $f \in \tilde{\mathsf{W}}$,

$$a_N^2 \tilde{\Omega}_N^{-2} \sum_{i=1}^{\alpha M_N} \tilde{\omega}_{N,i}^2 f(\tilde{\xi}_{N,i}) \xrightarrow{\mathbb{P}} \tilde{\gamma}(f). \tag{49}$$

However, since $\mathcal{R}_{0,k}(\cdot, \Phi_k^2 f) \leq \mathbb{1}_{\{\cdot:|f(\cdot)|>1\}}\mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2)+\mathbb{1}_{\{\cdot:|f(\cdot)|\leq 1\}}\mathcal{R}_{0,k}(\cdot, \Phi_k^2) \in \mathsf{W}$, a direct application of Douc and Moulines (2008, Equation (39)) yields that, for any $k \in \{1, \ldots, \alpha\}$,

$$a_N^2 \Omega_N^{-2} \sum_{j=1}^{M_N} \tilde{\omega}_{N,\alpha(j-1)+k}^2 f(\tilde{\xi}_{N,\alpha(j-1)+k}) \xrightarrow{\mathbb{P}} \gamma \mathcal{R}_{0,k}(\Phi_k^2 f).$$

Combining (49) with the limit $\tilde{\Omega}_N/(\alpha\Omega_N) \xrightarrow{\mathbb{P}} \nu L(\tilde{\Xi})$ [see (39)] we obtain, using Slutsky's theorem,

$$\begin{aligned} a_N^2 \tilde{\Omega}_N^{-2} \sum_{i=1}^{\alpha M_N} \tilde{\omega}_{N,i}^2 f(\tilde{\xi}_{N,i}) &= (\alpha\Omega_N/\tilde{\Omega}_N)^2 \alpha^{-2} \sum_{k=1}^{\alpha} a_N^2 \Omega_N^{-2} \sum_{j=1}^{M_N} \tilde{\omega}_{N,\alpha(j-1)+k}^2 f(\tilde{\xi}_{N,\alpha(j-1)+k}) \\ &\xrightarrow{\mathbb{P}} \alpha^{-2} \sum_{k=1}^{\alpha} \gamma \mathcal{R}_{0,k}(\Phi_k^2 f)/[\nu L(\tilde{\Xi})]^2 = \tilde{\gamma}(f). \end{aligned}$$

Finally, we establish the last property of Definition 3, that is,

$$a_N \tilde{\Omega}_N^{-1} \max_{1 \leq i \leq \alpha M_N} \tilde{\omega}_{N,i} \xrightarrow{\mathbb{P}} 0. \tag{50}$$

However, since, as shown by Douc and Moulines (2008, p. 30), for any $k \in \{1, \ldots, \alpha\}$,

$$a_N^2 (\alpha\Omega_N)^{-2} \max_{1 \leq j \leq M_N} \tilde{\omega}_{N,\alpha(j-1)+k}^2 \xrightarrow{\mathbb{P}} 0,$$

we immediately obtain

$$a_N^2 \tilde{\Omega}_N^{-2} \max_{1 \leq i \leq \alpha M_N} \tilde{\omega}_{N,i}^2 \leq (\alpha\Omega_N/\tilde{\Omega}_N)^2 \sum_{k=1}^{\alpha} a_N^2 (\alpha\Omega_N)^{-2} \max_{1 \leq j \leq M_N} \tilde{\omega}_{N,\alpha(j-1)+k}^2 \xrightarrow{\mathbb{P}} 0,$$

from which (50) follows.

It remains to show that the sets $\tilde{\mathsf{A}}$ and $\tilde{\mathsf{W}}$ are proper. Since, by assumption, $L(\cdot, \tilde{\Xi}) \in \mathsf{A}$ and $\mathcal{R}_{0,k}(\cdot, \Phi_k^2) \in \mathsf{W}$, $k \in \{1, \ldots, \alpha\}$, we conclude immediately that all constant functions $f \equiv c$ belong to $\tilde{\mathsf{A}}$. Now, let $|f| \leq |g|$, where g belongs to $\tilde{\mathsf{A}}$. Then $L(\cdot, |f|) \leq L(\cdot, |g|) \in \mathsf{A}$ and $\mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2) \leq \mathcal{R}_{0,k}(\cdot, \Phi_k^2 g^2) \in \mathsf{W}$, $k \in \{1, \ldots, \alpha\}$, implying, by property (ii) in the definition of a proper set, that $f \in \tilde{\mathsf{A}}$. Finally, let f and

g be any two functions in $\tilde{\mathsf{A}}$. Then, for any constants $(a, b) \in \mathbb{R}^2$, $L(\cdot, |af + bg|) \leq |a|L(\cdot, |f|) + |b|L(\cdot, |g|) \in \mathsf{A}$; moreover, for all $k \in \{1, \ldots, \alpha\}$,

$$\mathcal{R}_{0,k}(\cdot, \Phi_k^2[af + bg]^2) \leq (a^2 + |a|)\mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2) + (b^2 + |b|)\mathcal{R}_{0,k}(\cdot, \Phi_k^2 g^2) \in \mathsf{W},$$

implying that $af + bg \in \tilde{\mathsf{A}}$. The properness of $\tilde{\mathsf{W}}$ is established in a similar manner. This completes the proof. □

A.3 Demonstration of (36)

Since $U_i^{(n)}$ and $\epsilon_i^{(n)}$ are independent, it holds that

$$\begin{aligned}
&\operatorname{Cov}\left[\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1}, \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+2} \mid \xi^{(n)}_{N,i} = \xi\right] \\
&= \left[(\tau_n^{(1)}(\xi))^2 + (\tau_n^{(2)}(\xi))^2\right] \operatorname{Cov}\left[\mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi)\}}, \mathbb{1}_{\{1-U_i^{(n)} < \bar{\beta}_n(\xi)\}} \mid \xi^{(n)}_{N,i} = \xi\right] \\
&\quad + 2\tau_n^{(1)}(\xi)\tau_n^{(2)}(\xi) \operatorname{Cov}\left[\mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi)\}}, \mathbb{1}_{\{1-U_i^{(n)} \geq \bar{\beta}_n(\xi)\}} \mid \xi^{(n)}_{N,i} = \xi\right] - \eta_n^2.
\end{aligned} \tag{51}$$

In addition, as $U_i^{(n)}$ is independent of $\xi^{(n)}_{N,i}$ we obtain

$$\begin{aligned}
&\operatorname{Cov}\left[\mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi)\}}, \mathbb{1}_{\{1-U_i^{(n)} < \bar{\beta}_n(\xi)\}} \mid \xi^{(n)}_{N,i} = \xi\right] \\
&= \mathbb{P}\left(1 - \bar{\beta}_n(\xi) < U_i^{(n)} < \bar{\beta}_n(\xi) \mid \xi^{(n)}_{N,i} = \xi\right) - \bar{\beta}_n^2(\xi) \\
&= \mathbb{1}_{\{\bar{\beta}_n(\xi) > 1/2\}}(2\bar{\beta}_n(\xi) - 1) - \bar{\beta}_n^2(\xi)
\end{aligned} \tag{52}$$

and, analogously,

$$\begin{aligned}
&\operatorname{Cov}\left[\mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi)\}}, \mathbb{1}_{\{1-U_i^{(n)} \geq \bar{\beta}_n(\xi)\}} \mid \xi^{(n)}_{N,i} = \xi\right] \\
&= \mathbb{P}\left(U_i^{(n)} \leq \min\{\bar{\beta}_n(\xi), 1 - \bar{\beta}_n(\xi)\} \mid \xi^{(n)}_{N,i} = \xi\right) - \bar{\beta}_n(\xi)(1 - \bar{\beta}_n(\xi)) \\
&= \mathbb{1}_{\{\bar{\beta}_n(\xi) \leq 1/2\}}\bar{\beta}_n(\xi) + \mathbb{1}_{\{\bar{\beta}_n(\xi) > 1/2\}}(1 - \bar{\beta}_n(\xi)) - \bar{\beta}_n(\xi)(1 - \bar{\beta}_n(\xi)).
\end{aligned} \tag{53}$$

Now, assume that $\bar{\beta}_n(\xi) > 1/2$; then, using (51)–(53),

$$\begin{aligned}
&\operatorname{Cov}\left[\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1}, \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+2} \mid \xi^{(n)}_{N,i} = \xi\right] \\
&= -\left[(\tau_n^{(1)}(\xi))^2 + (\tau_n^{(2)}(\xi))^2\right](1 - \bar{\beta}_n(\xi))^2 + 2\tau_n^{(1)}(\xi)\tau_n^{(2)}(\xi)(1 - \bar{\beta}_n(\xi))^2 - \eta_n^2 \\
&= -\left(\tau_n^{(1)}(\xi) - \tau_n^{(2)}(\xi)\right)^2 (1 - \bar{\beta}_n(\xi))^2 - \eta_n^2.
\end{aligned}$$

Moreover, assuming that $\bar{\beta}_n(\xi) \leq 1/2$ yields similarly

$$\operatorname{Cov}\left[\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1}, \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+2} \mid \xi^{(n)}_{N,i} = \xi\right] = -\left(\tau_n^{(1)}(\xi) - \tau_n^{(2)}(\xi)\right)^2 \bar{\beta}_n^2(\xi) - \eta_n^2.$$

Finally, since $\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1}$ and $\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+2}$ have, conditionally on $\xi^{(n)}_{N,i}$, the same marginal distributions, and

$$\begin{aligned}\operatorname{Var}\left[\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1} \mid \xi^{(n)}_{N,i} = \xi\right] &= \left(\tau_n^{(1)}(\xi) - \tau_n^{(2)}(\xi)\right)^2 \operatorname{Var}\left[\mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi)\}} \mid \xi^{(n)}_{N,i} = \xi\right] + \eta_n^2 \\ &= \left(\tau_n^{(1)}(\xi) - \tau_n^{(2)}(\xi)\right)^2 \bar{\beta}_n(\xi)(1 - \bar{\beta}_n(\xi)) + \eta_n^2,\end{aligned}$$

the identity (36) follows. □

Acknowledgments We thank the anonymous referee for insightful comments that improved the presentation of the paper.

References

Arvidsen, N. I., Johnsson, T. (1982). Variance reduction through negative correlation—A simulation study. *Journal of Statistical Computation and Simulation*, *15*, 119–127.

Bollerslev, T., Engle, R. F., Nelson, D. B. (1994). ARCH models. In R. F. Engle, D. McFadden (Eds.), *The handbook of econometrics* (4, pp. 2959–3038). Amsterdam: North-Holland.

Cappé, O., Moulines, É., Rydén, T. (2005). *Inference in hidden Markov models*. New York: Springer.

Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Annals of Statistics*, *32*, 2385–2411.

Cornebise, J., Moulines, E., Olsson, J. (2014). Adaptive sequential Monte Carlo by means of mixture of experts. *Statistics and Computing*, *24*, 317–337.

Craiu, V. R., Meng, X.-L. (2005). Multiprocess parallel antithetic coupling for backward and forward Markov chain Monte Carlo. *Annals of Statistics*, *33*, 661–697.

Del Moral, P. (2004). *Feyman–Kac formulae. Genealogical and interacting particle systems with applications*. New York: Springer.

Douc, R., Moulines, É. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Annals of Statistics*, *36*, 2344–2376.

Douc, R., Moulines, É., Olsson, J. (2008). Optimality of the auxiliary particle filter. *Probability and Mathematical Statistics*, *29*, 1–28.

Doucet, A., de Freitas, N., Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. New York: Springer.

Doucet, A., Godsill, S., Andrieu, C. (2000). On sequential Monte-Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, *10*, 197–208.

Doucet, A., Johansen, A. M. (2011). Particle filtering and smoothing: Fifteen years later. In D. Crisan, B. Rozovsky (Eds.), *Oxford handbook of nonlinear filtering*. Oxford: Oxford University Press.

Fearnhead, P. (1998). Sequential Monte Carlo methods in filter theory. PhD thesis, University of Oxford.

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

Kitagawa, G. (1987). Non-Gaussian state space modeling of nonstationary time series. *Journal of the American Statistical Association*, *82*, 1023–1063.

Künsch, H. R. (2013). Particle filters. *Bernoulli*, *19*, 1391–1403.

Pitt, M. K., Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, *87*, 493–499.