ANTITHETIC SAMPLING FOR SEQUENTIAL MONTE CARLO METHODS WITH APPLICATION TO STATE SPACE MODELS

SVETLANA BIZJAJEVA AND JIMMY OLSSON

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. A version of the standard auxiliary particle filter (Pitt and Shephard, 1999) is proposed where the particles are mutated blockwise in such a way that all particles within each block are, firstly, offspring of a common ancestor and, secondly, 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.

1. Introduction

Sequential Monte Carlo (SMC) *methods*—alternatively termed *particle filters*—refer to a collection of algorithms which approximate recursively a sequence (often called the *Feynman-Kac flow*) of target measures by a sequence of empirical distributions associated with properly weighted samples of *particles.* These methods have received a lot of attention during the last decade and are at present applied within a wide range of scientific disciplines. Doucet *et al.* (2001) provides a survey of recent developments of the SMC methodology from a practical viewpoint and a comprehensive treatment of theoretical aspects of basic SMC algorithms is given by Del Moral (2004).

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 normalized 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) which relies on the classical idea of *antithetic sampling*

2000 *Mathematics Subject Classification.* Primary 65C05; Secondary 60G35.

Key words and phrases. Antithetic sampling, central limit theorem, optimal filtering, optimal kernel, particle filter, permuted displacement method, state space models.

used in standard Monte Carlo estimation: when estimating the expectation

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

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

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

of $I(f)$, where $\{\xi_i\}_{i=1}^N$ and $\{\xi_i'\}_{i=1}^N$ are two 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 the variables $f(\xi_i)$ and $f(\xi_i')$ are *negatively correlated* for all $i \in \{1, \ldots, N\}$. In this setting, the variables $\{\xi_i'\}_{i=1}^N$ are referred to as *antithetic variables*. Antithetically coupled variables can be generated in different ways, and in Section 2 we discuss how this can be achieved by means of the well-known *permuted displacement method* (Arvidsen and Johnsson, 1982). In order to allow for antithetic acceleration within the SMC framework we introduce (in Section 2) a version of the standard APF where the particles are mutated *blockwise* in such a way that all particles within each block are, firstly, offspring of a common ancestor and, secondly, statistically dependent conditionally on this ancestor. Moreover, in Section 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 3.2 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, Section 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, *besides* a lowered computational burden, a significant gain of precision for these models.

2. Auxiliary particle filter with blockwise correlated mutation

2.1. **Notation and definitions.** In order 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 $\mathcal{P}(\Xi)$ and $\mathbb{B}(\Xi)$ the sets of probability measures on $(\Xi, \mathcal{B}(\Xi))$ and measurable functions from Ξ to $\mathbb{R}$, respectively. For any measure $\mu \in \mathcal{P}(\Xi)$ and function $f \in \mathbb{B}(\Xi)$ satisfying $\int_{\Xi} |f(\xi)| \, \mu(\mathrm{d}\xi) < \infty$ we let $\mu(f) \triangleq \int_{\Xi} f(\xi) \, \mu(\mathrm{d}\xi)$ denote the expectation of f under μ. A 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 $K(\xi, \tilde{\Xi}) = 1$ for all $\xi \in \Xi$.

Moreover, a kernel K induces two operators, the first transforming a function $f \in \mathbb{B}(\Xi \times \tilde{\Xi})$ satisfying $\int_{\tilde{\Xi}} |f(\xi, \tilde{\xi})| \, K(\xi, \mathrm{d}\tilde{\xi}) < \infty$ into the function

$$\xi \mapsto K(\xi, f) \triangleq \int_{\tilde{\Xi}} f(\xi, \tilde{\xi}) \, K(\xi, \mathrm{d}\tilde{\xi})$$

in $\mathbb{B}(\Xi)$; the other transforms any measure $\nu \in \mathcal{P}(\Xi)$ into the measure

$$A \mapsto \nu K(A) \triangleq \int_{\Xi} K(\xi, A) \, \nu(\mathrm{d}\xi)$$

in $\mathcal{P}(\tilde{\Xi})$. Finally, in order 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 from $(\Xi, \mathcal{B}(\Xi))$ to the product space $\tilde{\Xi} \times \bar{\Xi}$, equipped with the product σ-algebra $\mathcal{B}(\tilde{\Xi}) \otimes \mathcal{B}(\bar{\Xi})$, given by

$$K \otimes T(\xi, A) \triangleq \iint_{\tilde{\Xi} \times \bar{\Xi}} \mathbb{1}_A(\tilde{\xi}, \bar{\xi}) \, K(\xi, \mathrm{d}\tilde{\xi}) \, T(\xi, \tilde{\xi}, \mathrm{d}\bar{\xi}) \;, \quad \xi \in \Xi \;, \; A \in \mathcal{B}(\tilde{\Xi}) \otimes \mathcal{B}(\bar{\Xi}) \;. \tag{2.1}$$

2.2. **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 $\nu \in \mathcal{P}(\Xi)$ if, denoting the weight sum by $\Omega_N \triangleq \sum_{i=1}^{M_N} \omega_{N,i}$,

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

for all functions f in some specified subset of $\mathbb{B}(\Xi)$. 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 $\nu \in \mathcal{P}(\Xi)$ 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

$$\mu(A) = \frac{\nu L(A)}{\nu L(\tilde{\Xi})} = \frac{\int_{\Xi} L(\xi, A) \, \nu(\mathrm{d}\xi)}{\int_{\Xi} L(\xi', \tilde{\Xi}) \, \nu(\mathrm{d}\xi')} \;, \quad A \in \mathcal{B}(\tilde{\Xi}) \;, \tag{2.2}$$

where L is a finite transition kernel from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. Feynman-Kac transitions of type (2.2) occur within a variety of fields (see Del Moral, 2004, for examples from, *e.g.*, quantum physics and biology) and in Section 4 we show how the flow of posterior distributions of the noisily observed Markov chain (state signal) of a state space model can be generated according to (2.2). The transformation is carried out by, firstly, 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 $A \in \mathcal{B}(\tilde{\Xi})$,

$$\mathbb{P}\left(\tilde{\xi}_{N,\alpha(j-1)+k} \in A \middle| \mathcal{F}_{N,\alpha(j-1)+k-1} \right) = R_k(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+1}, \ldots, \tilde{\xi}_{N,\alpha(j-1)+k-1}, A) \;,$$

where we have defined the σ-fields $\mathcal{F}_{N,\ell} \triangleq \sigma(\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}, \{\tilde{\xi}_{N,j}\}_{j=1}^{\ell})$, $\ell \in \{0, \ldots, \alpha M_N\}$, and each R_k is a Markovian kernel from $(\Xi \times \tilde{\Xi}^{k-1}, \mathcal{B}(\Xi \times \tilde{\Xi}^{k-1}))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. Hence, using the kernel outer product notation $\otimes$ defined in (2.1), the joint distribution, conditional on $\mathcal{F}_{N,\alpha(j-1)}$, the each block $\{\tilde{\xi}_{N,\alpha(j-1)+k}\}_{k=1}^{\alpha}$ can be expressed as $\bigotimes_{k=1}^{\alpha} R_k(\xi_{N,I_{N,j}}, \cdot)$. Secondly, these particles are associated with the weights

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

with

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

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

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

where we have introduced vector notation $a_{m:n} \triangleq (a_m, a_{m+1}, \ldots, a_n)$ 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 (2005), is summarized in Algorithm 1.

Algorithm 1 Blockwise correlated mutation

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

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

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

Here the mutation step (2) is expressed using the kernel outer product notation $\otimes$ defined in (2.1).

2.3. **Blockwise correlated mutation with resampling.** In the sequential context, where the problem consists in estimating a *sequence* of measures generated according to the mapping (2.2), it is, in order to avoid *weight degeneracy*, 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.

Algorithm 2 APF with blockwise correlated mutation

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

2.4. **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 $\mathcal{L}(\xi, A) \approx L(\xi, A)$ for all $\xi \in \Xi$ and $A \in \mathcal{B}(\tilde{\Xi})$; 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 $i \in \{1, \dots, M_N\}$ and $k \in \{1, \dots, \alpha\}$ returns, since then

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

a close to uniformly weighted particle sample. Thus, methods for approximating the optimal kernel have been proposed by several authors; see *e.g.* Pitt and Shephard (1999) and Doucet *et al.* (2000).

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 the variance of the produced SMC estimates significantly. Hence, the aim of this paper is to justify, in theory as well as in simulations, the following algorithm in which $\mathcal{L}$ and f denote a given approximation of L and a given target function, respectively.

Algorithm 3 APF with antithetic blockwise mutation

Require: $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ targets ν.

1: Draw $\{I_{N,j}\}_{j=1}^{\tilde{M}_N} \sim \mathcal{M}(\tilde{M}_N, \{\omega_{N,i}\mathcal{L}(\xi_{N,i}, \tilde{\Xi}) / \sum_{\ell=1}^{M_N} \omega_{N,\ell}\mathcal{L}(\xi_{N,\ell}, \tilde{\Xi})\}_{i=1}^{M_N})$,

2: **for** $j = 1, \ldots, \tilde{M}_N$ **do**

3: simulate, using an appropriate family of kernels $\{R_k\}_{k=1}^{\alpha}$, a block $\{\tilde{\xi}_{N,\alpha(j-1)+k}\}_{k=1}^{\alpha} \sim \bigotimes_{k=1}^{\alpha} R_k(\xi_{N,I_{N,j}}, \cdot)$ of particles such 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,

4: set, for $k \in \{1, \ldots, \alpha\}$,

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

5: **end for**

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

Step (3) in Algorithm 3 can be carried out in several different ways. 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 antithetis* (EA) *property*: if F is an arbitrary distribution function, 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 constraint that $\xi, \xi' \sim F$. This implies immediately that the strategy also achieves EA for variates $g(\xi)$ and $g(\xi')$, where $g : \mathbb{R} \to \mathbb{R}$ is any monotone function such that $\int g^2(\xi)\, F(\mathrm{d}\xi) < \infty$, since (U, U') achieves EA simultaneously for all F and $g(\xi)$ (and $g(\xi')$) has distribution function $F \circ g^{\leftarrow}$. 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 2.1 (Pairwise negative association)**.** *The random variables $\xi_1, \xi_2, \ldots, \xi_n$ 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 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 $\{1, \dots, \alpha\}$.

Algorithm 4 Permuted displacement method

Draw $r_1 \sim \mathcal{U}(0,1)$,
for $k = 2, \dots, \alpha - 1$ **do**
  set $r_k = \langle 2^{k-2} r_1 + 1/2 \rangle$,
end for
set $r_\alpha = 1 - \langle 2^{\alpha-2} r_1 \rangle$,
pick a random $\sigma \in S_\alpha$,
for $k = 1, \dots, \alpha$ **do**
  set $U_k \triangleq r_{\sigma(k)}$,
end for

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 (3) of Algorithm 3 can be carried out by producing, using Algorithm 4, PNA uniforms $\{U_k\}_{k=1}^{\alpha}$ and setting, for $k \in \{1, \dots, \alpha\}$,

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

where $F_{k,\xi}[f](x) \triangleq \mathcal{L}(\xi, \{f(\tilde{\xi}) \leq x\})/\mathcal{L}(\xi, \tilde{\Xi})$, $x \in \mathbb{R}$, denotes the conditional distribution function of the $f(\tilde{\xi}_{N,\alpha(j-1)+k})$'s given $\xi_{N,j} = \xi \in \Xi$. Since each function $F^{\leftarrow}_{k,\xi}[f]$ is monotone, it follows that $\{f(\tilde{\xi}_{N,\alpha(j-1)+k})\}_{k=1}^{\alpha}$ are conditionally EA. Of course, this method is applicable only when $F_{k,\xi}[f]$ is easy to invert; this is however not always the case and in Section 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 novel results on triangular arrays obtained by Douc and Moulines (2005). The arguments rely on results describing the weak convergence of Algorithms 1 and 2 in a rather general setting.

3.1. **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 (2005), 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 3.1 (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* $N \to \infty$*,*

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

Definition 3.2 (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* $N \to \infty$*,*

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

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

$$\begin{aligned}
\tilde{\mathsf{C}} &\triangleq \{f \in \mathsf{L}^1(\tilde{\Xi}, \mu) : L(\cdot, |f|) \in \mathsf{C}\} \; , \\
\tilde{\mathsf{A}} &\triangleq \left\{f : L(\cdot, |f|) \in \mathsf{A}, \mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2) \in \mathsf{W}; k \in \{1, \ldots, \alpha\}\right\} \; , \\
\tilde{\mathsf{W}} &\triangleq \left\{f : \mathcal{R}_{0,k}(\cdot, \Phi_k^2 |f|) \in \mathsf{W}; k \in \{1, \ldots, \alpha\}\right\} \; .
\end{aligned} \tag{3.1}$$

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

$$\begin{aligned}
&\mathbb{M}_{m,n}(\xi, f) \\
&\triangleq \mathbb{E}\left[\Phi_m(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+m}) \Phi_n(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+n}) f(\tilde{\xi}_{N,\alpha(j-1)+m}) f(\tilde{\xi}_{N,\alpha(j-1)+n}) \middle| \xi_{N,j} = \xi\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\Phi_n(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+n}) f(\tilde{\xi}_{N,\alpha(j-1)+n}) \middle| \xi_{N,j} = \xi, \tilde{\xi}_{N,\alpha(j-1)+1:\alpha(j-1)+m}\right]\right. \\
&\qquad\qquad \left. \times \Phi_m(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+m}) f(\tilde{\xi}_{N,\alpha(j-1)+m}) \middle| \xi_{N,j} = \xi\right] \\
&= \int_{\tilde{\Xi}} \cdots \int_{\tilde{\Xi}} \mathcal{R}_{m,n}(\xi, \tilde{\xi}_{1:m}, \Phi_n(\xi, \cdot) f) \Phi_m(\xi, \tilde{\xi}_m) f(\tilde{\xi}_m) \bigotimes_{\ell=1}^{m} R_\ell(\xi, \mathrm{d}\tilde{\xi}_1 \times \cdots \times \mathrm{d}\tilde{\xi}_m) \; ,
\end{aligned}$$

and introduce the conditional covariances

$$\begin{aligned}&\mathbb{C}_{m,n}(\xi, f) \\ &\triangleq \operatorname{Cov}\left[\Phi_m(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+m}) f(\tilde{\xi}_{N,\alpha(j-1)+m}), \Phi_n(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+n}) f(\tilde{\xi}_{N,\alpha(j-1)+n}) \middle| \xi_{N,j} = \xi\right] \\ &\qquad = \mathbb{M}_{m,n}(\xi, f) - L^2(\xi, f) \; . \quad (3.2)\end{aligned}$$

3.2. **Convergence of Algorithms 1 and 2.** Under the assumptions above we have the following convergence results, whose proofs are found in the appendix.

Theorem 3.1. *Assume* **(A1)** *and suppose that* $L(\cdot, \tilde{\Xi}) \in \mathsf{C}$. *Then the set* $\tilde{\mathsf{C}}$ *defined in* (3.1) *is proper and the weighted sample* $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha M_N}$ *produced in Algorithm 1 is consistent for* $(\mu, \tilde{\mathsf{C}})$.

Theorem 3.2. *Let the assumptions of Theorem 3.1 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* (3.1) *are proper and the weighted sample* $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha M_N}$ *produced in Algorithm 1 is AN for* $(\mu, \tilde{\mathsf{A}}, \tilde{\mathsf{W}}, \tilde{\sigma}, \tilde{\gamma}, \{a_N\}_{N=1}^{\infty})$, *where, for* $f \in \tilde{\mathsf{A}}$,

$$\tilde{\sigma}^2(f) \triangleq \sigma^2\{L[f - \mu(f)]\}/[\nu L(\tilde{\Xi})]^2 + \sum_{(m,n)\in\{1,\ldots,\alpha\}^2} \gamma \mathbb{C}_{m,n}[f - \mu(f)]/[\alpha \nu L(\tilde{\Xi})]^2 \; , \qquad (3.3)$$

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

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

Remark 3.1. *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* (3.3) *of form*

$$\begin{aligned}\tilde{\sigma}^2(f) &= \sigma^2\{L[f - \mu(f)]\}/[\nu L(\tilde{\Xi})]^2 + \sum_{m=1}^{\alpha} \gamma \mathbb{C}_{m,m}[f - \mu(f)]/[\alpha \nu L(\tilde{\Xi})]^2 \\ &= \sigma^2\{L[f - \mu(f)]\}/[\nu L(\tilde{\Xi})]^2 + \alpha^{-1}\{\gamma R(\Phi^2[f - \mu(f)]^2) - \gamma L^2[f - \mu(f)]\}/[\nu L(\tilde{\Xi})]^2 \; , \qquad (3.4)\end{aligned}$$

which is exactly the expression obtained by Douc and Moulines (2005, 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

$$
\begin{aligned}
\bar{\mathsf{C}} &\triangleq \{f \in \mathsf{L}^1(\mu, \tilde{\Xi}) : L(\cdot, |f|) \in \mathsf{C} \cap \mathsf{L}^1(\nu, \tilde{\Xi})\} , \\
\bar{\mathsf{A}} &\triangleq \big\{\Psi^{-1} L^2(\cdot, |f|) \in \mathsf{C} \cap \mathsf{L}^1(\nu, \Xi), L(\cdot, |f|) \in \mathsf{A}, L^2(\cdot, |f|) \in \mathsf{W}, \\
&\qquad\qquad \Psi^{-1} \mathcal{R}_{0,k}(\cdot, \Phi_k^2 f^2) \in \mathsf{C} \cap \mathsf{L}^1(\nu, \Xi); k \in \{1, \ldots, \alpha\}\big\} , \\
\bar{\mathsf{W}} &\triangleq \big\{\Psi^{-1} \mathcal{R}_{0,k}(\cdot, \Phi_k^2 |f|) \in \mathsf{C} \cap \mathsf{L}^1(\nu, \Xi); k \in \{1, \ldots, \alpha\}\big\} ;
\end{aligned}
\tag{3.5}
$$

now, by combining Theorem 3.2 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 are omitted for brevity.

Corollary 3.1. *Let the assumptions of Theorem 3.1 hold and assume* **(A3)**. *Then the set* $\bar{\mathsf{C}}$ *defined in* (3.5) *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 3.2. *Let the assumptions of Theorem 3.1 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* (3.5) *are proper and the weighted sample* $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\alpha \tilde{M}_N}$ *obtained in 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* $f \in \bar{\mathsf{A}}$,

$$
\begin{aligned}
\bar{\sigma}^2[\Psi](f) &\triangleq \sigma^2\{L[\cdot, f - \mu(f)]\} / [\nu L(\tilde{\Xi})]^2 \\
&\quad + \beta \ell^{-1} \nu(\Psi) \sum_{(m,n) \in \{1, \ldots, \alpha\}^2} \nu(\Psi \mathbb{M}_{m,n}\{\cdot, \Psi^{-1}[f - \mu(f)]\}) / [\alpha \nu L(\tilde{\Xi})]^2
\end{aligned}
\tag{3.6}
$$

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

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

Remark 3.2. *The resampling step (1) in Algorithm 2 can, of course, be based on resampling techniques different from multinomial resampling,* e.g., residual resampling *or* Bernoulli branching. *However, we believe that the convergence results stated in Theorems 3.1 and 3.2 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, 2005).*

3.3. **Theoretical justification of Algorithm 3.** In order to justify the use of antithetic variables in Algorithm 3, we examine the asymptotic variance given in (3.6). Since the first term is not at all effected by the way the particles are mutated, we direct focus to the second term and write, using (3.2),

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

where the first term on the RHS is again independent of the correlation structure of the mutation step. The second term will be smaller than in the case where all particles within each block are mutated independently if the covariances $\mathbb{C}_{m,n}\{\cdot, \Psi^{-1}[f-\mu(f)]\}$ are negative for all $m \neq n$; however, since $\Psi^{-1}(\xi)\Phi(\xi,\tilde{\xi}) \approx 1$ for all $(\xi,\tilde{\xi}) \in \Xi \times \tilde{\Xi}$ in the close to fully adapted case, it holds that

$$\mathbb{C}_{m,n}\{\xi, \Psi^{-1}[f-\mu(f)]\} \approx \operatorname{Cov}\left[f(\tilde{\xi}_{N,\alpha(j-1)+m}), f(\tilde{\xi}_{N,\alpha(j-1)+n}) \middle| \xi_{N,j} = \xi \right] , \tag{3.7}$$

which is negative when the $f(\tilde{\xi}_{N,\alpha(j-1)+k})$'s are negatively correlated.

In addition, it is possible to relate the performance of the antithetic SMC scheme in Algorithm 3 to that of the standard APF (for which $\alpha = 1$). More specifically, we establish a criterion (depending on the model and target function under consideration) which guarantees that introducing antithetic variates yields a strictly more accurate (in terms of variance) *and* computationally more efficient algorithm than the standard APF. In order to keep the particle population size constant, *i.e.* having $\tilde{M}_N = M_N$, through a run of Algorithm 3 for a given block size α, only a fraction $\tilde{M}_N = \lceil M_N/\alpha \rceil$ (yielding $\ell = 1/\alpha$ in Corollary 3.2) of the original particle population should be selected at the resampling operation. In this case Corollary 3.2 provides, using (3.2), the asymptotic variance

$$\begin{aligned}\bar{\sigma}^2[\Psi](f) \triangleq{}& \sigma^2\{L[\cdot, f-\mu(f)]\}/[\nu L(\tilde{\Xi})]^2 \\ &+ \beta\alpha\nu(\Psi) \sum_{(m,n)\in\{1,\dots,\alpha\}^2} \nu(\Psi\mathbb{M}_{m,n}\{\cdot, \Psi^{-1}[f-\mu(f)]\})/[\alpha\nu L(\tilde{\Xi})]^2 .\end{aligned} \tag{3.8}$$

On the other hand, letting $\alpha = 1$ and $\ell = 1$, corresponding to the uncorrelated standard APF, in Corollary 3.2 yields the asymptotic variance

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

and, under the assumption that the inherent covariance structure of each block is *uniform* with $\mathbb{M}_{m,n} = \mathbb{M}^*$ for all $(m,n) \in \{1,\dots,\alpha\}^2$ such that $m \neq n$, the citerion

$$\begin{gathered}\bar{\sigma}^2[\Psi](f) \leq \bar{\sigma}^2_{\alpha=\ell=1}[\Psi](f) \\ \Leftrightarrow \\ -\nu(\Psi\mathbb{C}^*\{\cdot, \Psi^{-1}[f-\mu(f)]\}) \geq \nu(\Psi L^2\{\cdot, \Psi^{-1}[f-\mu(f)]\}) .\end{gathered} \tag{3.9}$$

Remark 3.3. *From the criterion* (3.9) *it is evident that mutating the particles in blocks without any (or positive) inherent correlation structure (that is, letting* $\mathbb{C}^* \geq 0$*) will, not surprisingly, increase the asymptotic variance vis-à-vis the standard APF. Moreover, since the correlation* $\mathbb{C}^* \geq 0$ *is a decreasing function of* α*, we conclude that there is a critical block size above which* (3.9) *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 modeled 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 Y_n depends on X_n only. For a model of this type, all inference about the hidden states has to be made through 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,

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

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

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

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

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

A straightforward application of Bayes's rule yields, for $A \in \mathcal{B}(\mathsf{X})$, the recursion

$$\begin{aligned} \phi_0(A) &= \frac{\int_A g_0(x, Y_0)\, \nu_0(\mathrm{d}x)}{\int_{\mathsf{X}} g_0(x, Y_0)\, \nu_0(\mathrm{d}x)} \ , \\ \phi_{n+1}(A) &= \frac{\int_{\mathsf{X}} \int_A g_{n+1}(x', Y_{n+1})\, Q_n(x, \mathrm{d}x')\, \phi_n(\mathrm{d}x)}{\iint_{\mathsf{X}^2} g_{n+1}(x', Y_{n+1})\, Q_n(x, \mathrm{d}x')\, \phi_n(\mathrm{d}x)} \ , \end{aligned} \tag{4.1}$$

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 apply the SMC methodology described in the previous; indeed, having defined, for $A \in \mathcal{B}(\mathsf{X})$ and $x \in \mathsf{X}$, the unnormalized transition kernels

$$L_n(x, A) = \int_A g_{n+1}(x', Y_{n+1})\, Q_n(x, \mathrm{d}x') \;, \tag{4.2}$$

yielding the equivalent Feynman-Kac representation

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

of (4.1), we conclude that the optimal filtering problem can be perfectly cast into the framework of Section 2 with $\Xi = \tilde{\Xi} = \mathsf{X}$, $\nu = \phi_n$, $L = L_n$, and $\mu = \phi_{n+1}$.

4.1. **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 normal 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 on closed form as

$$\Psi_n(x) = \mathcal{N}(Y_{n+1}; 0, \beta_0 + \beta_1 x^2 + \sigma^2) \tag{4.3}$$

where $\mathcal{N}(x; \mu, \sigma^2) \triangleq \exp[-(x-\mu)^2/(2\sigma^2)]/\sqrt{2\pi\sigma^2}$ denotes 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 $\{\phi_n(\mathrm{I}_{\mathsf{X}})\}_{n=0}^{30}$, where I_{X} denotes the identity mapping $\mathrm{I}_{\mathsf{X}}(x) = x$ 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

$$\begin{aligned} \tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)} &= m_n(\xi_{N,i}^{(n)}) + \hat{\sigma}_n(\xi_{N,i}^{(n)})\epsilon_i^{(n)} \;, \\ \tilde{\xi}_{N,\alpha(i-1)+2}^{(n+1)} &= 2m_n(\xi_{N,i}^{(n)}) - \tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)} \;, \end{aligned} \tag{4.4}$$

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

$$\begin{aligned}
\tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)} &= m_n(\xi_{N,i}^{(n)}) + \hat{\sigma}_n(\xi_{N,i}^{(n)})\epsilon_{i,1}^{(n)} , \\
\tilde{\xi}_{N,\alpha(i-1)+2}^{(n+1)} &= \frac{1}{2}\left(3m_n(\xi_{N,i}^{(n)}) - \tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)} + \sqrt{3}\hat{\sigma}_n(\xi_{N,i}^{(n)})\epsilon_{i,2}^{(n)}\right) , \\
\tilde{\xi}_{N,\alpha(i-1)+3}^{(n+1)} &= 3m_n(\xi_{N,i}^{(n)}) - \tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)} - \tilde{\xi}_{N,\alpha(i-1)+2}^{(n+1)} ,
\end{aligned} \tag{4.5}$$

where the independent sequences $\{\epsilon_{i,1}^{(n)}\}_{i=1}^{M_N}$ and $\{\epsilon_{i,2}^{(n)}\}_{i=1}^{M_N}$ are as above. The coupling (4.5) yields the conditional correlation $\mathrm{Corr}(\tilde{\xi}_{N,\alpha(i-1)+m}^{(n+1)}, \tilde{\xi}_{N,\alpha(i-1)+m'}^{(n+1)} | \xi_{N,i}^{(n)}) = -1/2$, 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 parametrized 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 $M_N = 6{,}000/\alpha$ are, for the different values of α, displayed in Figure 1(a) (the informative case) and Figure 1(b) (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 the variance significantly. Moreover, the improvement is especially noticeable in the informative case.

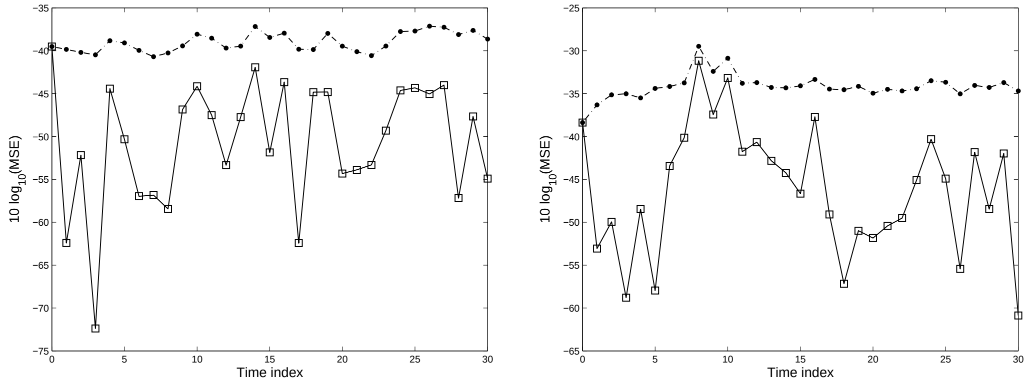
More relevant is to compare the performance of Algorithm 3, again with $\alpha \in \{2, 3\}$ and $M_N = 6{,}000/\alpha$, to that of the standard fully adapted APF using 6,000 particles without any block structure. In this setting, both antithetic filters are clearly more computationally efficient since, firstly, only a half and a third of the particles are selected at each resampling operation, and, secondly, 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 Figure from which it is clear that performances of the antithetic filters are, despite being less costly, superior, especially in the case of informative observations (Figure 2(a)); indeed, the improvement is over 20 Decibel at some time steps. Moreover, it is evident that the computational gain of using $\alpha = 3$ instead of $\alpha = 2$ offspring in each block is at the expense of a slight decrease of precision.

4.2. **Growth model.** The univariate *growth model* given by, for $n \geq 0$,

$$\begin{aligned}
X_{n+1} &= a_n(X_n) + \sigma_w W_{n+1} , \\
Y_n &= bX_n^2 + \sigma_v V_n ,
\end{aligned} \tag{4.6}$$

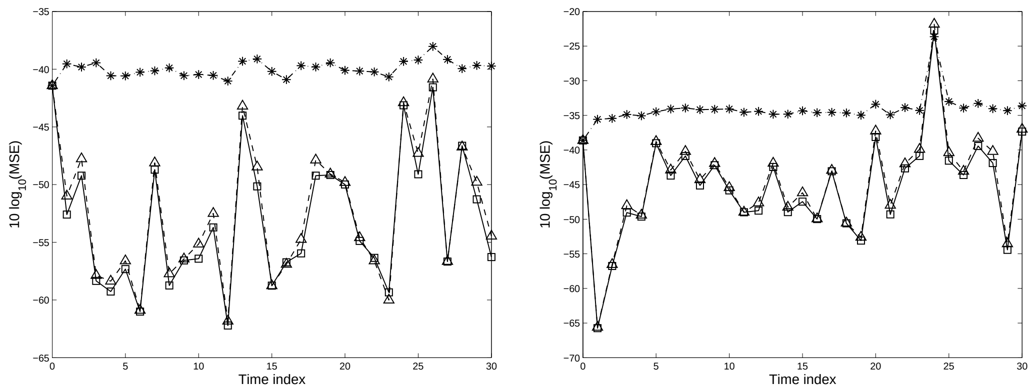
where

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



(a) (b)

Figure 1. Plot of MSEs (in Decibel) of filters being implementations of Algorithm 3 with $\alpha = 2$ antithetically coupled ($\square$) and conditionally independent ($\bullet$) 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 $M_N = 3{,}000$.



(a) (b)

Figure 2. Plot of MSEs (in Decibel) of the standard optimal APF ($*$) with 6,000 particles and antithetic filters with $\alpha = 2$ ($\square$) and $\alpha = 3$ ($\triangle$) for the ARCH model with informative (a) and non-informative (b) observations. $\alpha M_N = 6,000$ for both antithetic filters and the MSE values are based on 400 runs of each algorithm.

and the sequences $\{W_n\}_{n=1}^{\infty}$ and $\{W_n\}_{n=1}^{\infty}$ are as in the previous example, was discussed by Kitagawa (1987) (see also Polson *et al.*, 2002) 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

$$x \in \mathbb{R} \mapsto g(x, Y_n) = \mathcal{N}(Y_n; bx_n^2, \sigma_v^2) \in \mathbb{R}^+ \ . \tag{4.7}$$

which is symmetric about zero for any observation Y_n. Interestingly, functions (4.7) 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 equation (4.6) 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 (4.7) by a mixture

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

of two Gaussian densities, where

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

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 negated log curvature 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 (4.2) of the unnormalized optimal kernel yields straightforwardly the mixture

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

where each Gaussian stratum

$$G_n^{(d)}(x, A) \triangleq \int_A \mathcal{N}(x'; \tau_n^{(d)}(x), \eta_n^2) \, \mathrm{d}x' \ , \quad d \in \{1, 2\} \ ,$$

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

$$\tau_n^{(d)}(x) \triangleq \frac{\sigma_w^2 \mu_d + \varsigma^2 a_n(x)}{\sigma_w^2 + \varsigma^2} \ ,$$

$$\eta_n^2 \triangleq \frac{\sigma_w^2 \varsigma^2}{\sigma_w^2 + \varsigma^2} \ ,$$

is weighted by

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

By normalizing we obtain the approximation

$$\mathcal{L}_n(x, A)/\mathcal{L}_n(x, \mathsf{X}) = \bar{\beta}_n(x) G_n^{(1)}(x, A) + (1 - \bar{\beta}_n(x)) G_n^{(2)}(x, A) \ , \quad x \in \mathsf{X} \ , A \in \mathcal{B}(\mathsf{X}) \ , \tag{4.8}$$

of the optimal kernel, where we have defined the normalized 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 (4.8) as proposal, the experiment of the previous example (in which we estimated filter posterior means $\{\phi_n(\mathrm{I}_\mathsf{X})\}_{n=0}^{30}$) was repeated with focus set on the case $\alpha = 2$. In order impose a conditionally negative correlation structure we let each pair of offspring particles evolve according to

$$\begin{aligned} \tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)} &= \tau_n^{(1)}(\xi_{N,i}^{(n)}) \mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi_{N,i}^{(n)})\}} + \tau_n^{(2)}(\xi_{N,i}^{(n)}) \mathbb{1}_{\{U_i^{(n)} \geq \bar{\beta}_n(\xi_{N,i}^{(n)})\}} + \eta_n \epsilon_i^{(n)} \ , \\ \tilde{\xi}_{N,\alpha(i-1)+2}^{(n+1)} &= \tau_n^{(1)}(\xi_{N,i}^{(n)}) \mathbb{1}_{\{1-U_i^{(n)} < \bar{\beta}_n(\xi_{N,i}^{(n)})\}} + \tau_n^{(2)}(\xi_{N,i}^{(n)}) \mathbb{1}_{\{1-U_i^{(n)} \geq \bar{\beta}_n(\xi_{N,i}^{(n)})\}} - \eta_n \epsilon_i^{(n)} , \end{aligned} \tag{4.9}$$

where $\{U_i^{(n)}\}_{i=1}^{M_N}$ and $\{\epsilon_i^{(n)}\}_{i=1}^{M_N}$ are independent sequences of mutually independent uniformly (on $[0, 1]$) and standard normal distributed random variables, respectively, such that each pair $(U_i^{(n)}, \epsilon_i^{(n)})$ is independent of everything else. It is easily established 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 (4.9) is marginally distributied according to the approximate optimal kernel (4.8). In addition, one can show that (see Section A.3 for details) the correlation between the offspring of a block is given by, for $\xi \in \mathsf{X}$,

$$\begin{aligned} &\mathrm{Corr}\left[\tilde{\xi}_{N,\alpha(i-1)+1}^{(n+1)}, \tilde{\xi}_{N,\alpha(i-1)+2}^{(n+1)} \,\middle|\, \xi_{N,i}^{(n)} = \xi \right] \\ &= -\frac{(\tau_n^{(1)}(\xi) - \tau_n^{(2)}(\xi))^2 [\bar{\beta}_n^2(\xi) \mathbb{1}\{\bar{\beta}_n(\xi) \leq 1/2\} + (\bar{\beta}_n^2(\xi) - 1)^2 \mathbb{1}\{\bar{\beta}_n(\xi) > 1/2\}] + \eta_n^2}{(\tau_n^{(1)}(\xi) - \tau_n^{(2)}(\xi))^2 \bar{\beta}_n(\xi)(1 - \bar{\beta}_n(\xi)) + \eta_n^2} \ , \end{aligned} \tag{4.10}$$

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 $\alpha M_N = 5{,}000$, a (close to) fully adapted APF, based on the proposal kernel (4.8) and 5,000 particles, and the plain bootstrap filter using 5,000 particles. Like in the ARCH example, we let the filters approximate filter posterior means $\phi_n(\mathrm{I}_\mathsf{X})$ 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 ($\sigma_w^2 = 10$, Figure 3(a)) as well as non-informative ($\sigma_w^2 = 1$, Figure 3(b)) observations and the MSEs, measured with respect to reference values obtained with the 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

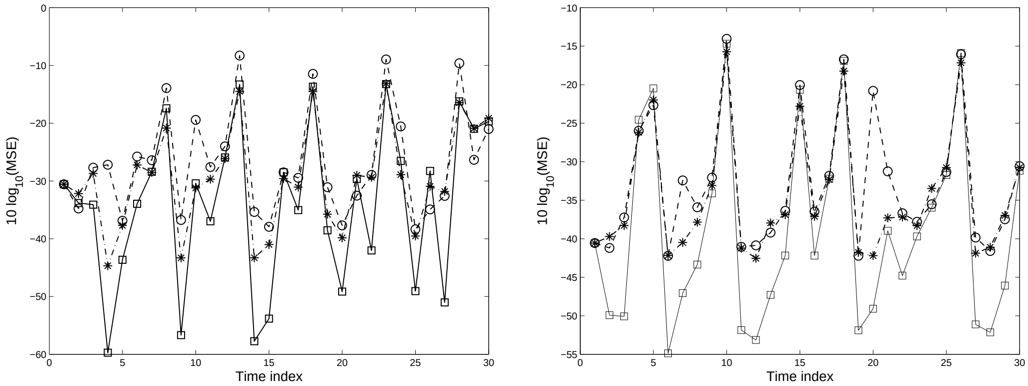


FIGURE 3. Plot of MSEs (in Decibel) of the plain bootstrap filter ($\circ$) using 5,000 particles, the standard optimal APF ($*$) using 5,000 particles, and the antithetic filter ($\Box$) with $\alpha = 2$ and $\alpha M_N = 5{,}000$ for the growth model with informative (a) and non-informative (b) observations. The MSE values are based on 400 runs of each algorithm.

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 (4.8) instead of the prior kernel yields, as we may expect, generally more precise posterior filter mean estimates, since the APF outperforms the bootstrap particle filter at most time steps.

APPENDIX A. PROOFS

A.1. **Proof of Theorem 3.1.** The result follows straightforwardly from Slutsky's theorem and results obtained by Douc and Moulines (2005) in the case of independently mutated particles. Indeed, by (Douc and Moulines, 2005, Equation (36)) we have, for any $1 \leq k \leq \alpha$,

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

yielding immediately

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

By applying (A.1) 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 3.1, write

$$(\alpha \Omega_N)^{-1} \max_{1 \leq i \leq \alpha M_N} \tilde{\omega}_{N,i} \leq \alpha^{-1} \sum_{k=1}^{\alpha} \tilde{\Omega}_N^{-1} \max_{1 \leq j \leq M_N} \tilde{\omega}_{N,\alpha(j-1)+k} \; ; \tag{A.2}$$

however, by inspecting the proof of (Douc and Moulines, 2005, Theorem 1) we conclude that each term on the RHS of (A.2) tends to zero in probability, which in combination with (A.1) implies that

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

This completes the proof.

A.2. **Proof of Theorem 3.2.** 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, 2005, Theorem 2),

$$a_N \tilde{\Omega}_N^{-1} \sum_{i=1}^{\alpha M_N} \tilde{\omega}_{N,i} f(\tilde{\xi}_{N,i}) = \alpha \Omega_N \tilde{\Omega}_N^{-1} (A_N + B_N) \; , \tag{A.3}$$

where

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

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 (A.1), $\tilde{\Omega}_N / (\alpha \Omega_N) \stackrel{\mathbb{P}}{\longrightarrow} \nu L(\tilde{\Xi})$, as $N \to \infty$, it is enough to prove that

$$A_N + B_N \stackrel{\mathcal{D}}{\longrightarrow} \mathcal{N}\{0, \sigma^2[L(\cdot, f)] + \eta^2(f)\} \; ,$$

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

$$A_N = a_N(\alpha\Omega_N)^{-1} \sum_{j=1}^{M_N} \sum_{k=1}^{\alpha} \mathbb{E}\left[\tilde{\omega}_{N,\alpha(j-1)+k} f(\tilde{\xi}_{N,\alpha(j-1)+k}) \middle| \mathcal{F}_{N,\alpha(j-1)} \right]$$
$$= a_N \Omega_N^{-1} \sum_{j=1}^{M_N} \omega_{N,j} L(\xi_{N,j}, f) \xrightarrow{\mathcal{D}} \mathcal{N}\{0, \sigma^2[L(\cdot, f)]\} ,$$

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

$$\mathbb{E}\left[\exp(\mathrm{i}uB_N) \middle| \mathcal{F}_{N,0}\right] \xrightarrow{\mathbb{P}} \exp(-u^2\eta^2(f)/2) \tag{A.4}$$

from which the result of the theorem follows. The proof of (A.4) consists in showing that the two conditions of Theorem 13 in (Douc and Moulines, 2005) 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

$$\mathbb{E}\left[U_{N,j}^2 \middle| \mathcal{F}_{N,\alpha(j-1)} \right]$$
$$= a_N^2(\alpha\Omega_N)^{-2} \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \mathbb{E}\left[\tilde{\omega}_{N,\alpha(j-1)+k} f(\tilde{\xi}_{N,\alpha(j-1)+k}) \tilde{\omega}_{N,\alpha(j-1)+m} f(\tilde{\xi}_{N,\alpha(j-1)+m}) \middle| \mathcal{F}_{N,\alpha(j-1)} \right]$$
$$= a_N^2(\alpha\Omega_N)^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \mathbb{M}_{k,m}(\xi_{N,j}, f) . \tag{A.5}$$

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

$$a_N^2(\alpha\Omega_N)^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \mathbb{M}_{k,m}(\xi_{N,j}, f) \xrightarrow{\mathbb{P}} \alpha^{-2} \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \gamma \mathbb{M}_{k,m}(f) . \tag{A.6}$$

Now consider

$$\sum_{j=1}^{M_N} \mathbb{E}^2\left[U_{N,j} \middle| \mathcal{F}_{N,\alpha(j-1)} \right]$$
$$= a_N^2(\alpha\Omega_N)^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \mathbb{E}^2\left[\sum_{k=1}^{\alpha} \Phi_k(\xi_{N,j}, \tilde{\xi}_{N,\alpha(j-1)+k}) f(\tilde{\xi}_{N,\alpha(j-1)+k}) \middle| \mathcal{F}_{N,\alpha(j-1)} \right]$$
$$= a_N^2 \Omega_N^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 L^2(\xi_{N,j}, f) ; \tag{A.7}$$

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

$$a_N^2 \Omega_N^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 L^2(\xi_{N,j}, f) \xrightarrow{\mathbb{P}} \gamma L^2(f) \ . \tag{A.8}$$

Finally, by combining Equations (A.5)–(A.8) we conclude that

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

which establishes condition *i)*.

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

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

Thus, argue along the lines of the proof of (Douc and Moulines, 2005, Theorem 2) and write, for any $C > 0$,

$$C_N \leq a_N^2 (\alpha \Omega_N)^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \mathbb{M}_{k,m}\left(\xi_{N,j}, f \mathbb{1}_{\{|\sum_{k=1}^{\alpha} \Phi_k f| \geq C\}} \right) \\ + \mathbb{1}_{\{a_N (\alpha \Omega_N)^{-1} \max_i \omega_{N,i} \geq \varepsilon C^{-1}\}} \sum_{j=1}^{M_N} \mathbb{E}\left[U_{N,j}^2 \,\middle|\, \mathcal{F}_{N,\alpha(j-1)} \right] . \tag{A.9}$$

Under **(A2)** the indicator function of the second term on the RHS of (A.9) 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

$$a_N^2 (\alpha \Omega_N)^{-2} \sum_{j=1}^{M_N} \omega_{N,j}^2 \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \mathbb{M}_{k,m}\left(\xi_{N,j}, f \mathbb{1}_{\{|\sum_{k=1}^{\alpha} \Phi_k f| \geq C\}} \right) \\ \xrightarrow{\mathbb{P}} \alpha^{-2} \sum_{(k,m)\in\{1,\ldots,\alpha\}^2} \gamma \mathbb{M}_{k,m}\left(f \mathbb{1}_{\{|\sum_{k=1}^{\alpha} \Phi_k f| \geq C\}} \right) . \tag{A.10}$$

By dominated convergence, the RHS of (A.10) can be made arbitrarily small by taking C sufficiently large. Therefore, also condition *ii)* is satisfied, implying the convergence (A.4). This establishes (A.3).

We turn to the second property of Definition 3.2 and show that, for any $f \in \tilde{\mathsf{W}}$,

$$a_N^2 \tilde{\Omega}_N^{-2} \sum_{i=1}^{\alpha M_N} \tilde{\omega}_{N,i}^2 f(\tilde{\xi}_{N,i}) \xrightarrow{\mathbb{P}} \tilde{\gamma}(f) . \tag{A.11}$$

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, 2005, 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 (A.11) with the limit $\tilde{\Omega}_N / (\alpha \Omega_N) \xrightarrow{\mathbb{P}} \nu L(\tilde{\Xi})$ (see (A.1)) 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 3.2, that is,

$$a_N \tilde{\Omega}_N^{-1} \max_{1 \leq i \leq \alpha M_N} \tilde{\omega}_{N,i} \xrightarrow{\mathbb{P}} 0 . \tag{A.12}$$

However, since, as shown by Douc and Moulines (2005, 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 (A.12) 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 (4.10).** Since $U_i^{(n)}$ and $\epsilon_i^{(n)}$ are independent, it holds that

$$\begin{aligned}
&\operatorname{Cov}\left[\tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+1}, \tilde{\xi}^{(n+1)}_{N,\alpha(i-1)+2} \,\middle|\, \xi^{(n)}_{N,i} = \xi\right] \\
&\quad = [(\tau_n^{(1)}(\xi))^2 + (\tau_n^{(2)}(\xi))^2] \operatorname{Cov}\left[\mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi)\}}, \mathbb{1}_{\{1-U_i^{(n)} < \bar{\beta}_n(\xi)\}} \,\middle|\, \xi^{(n)}_{N,i} = \xi\right] \\
&\qquad + 2\tau_n^{(1)}(\xi)\tau_n^{(2)}(\xi) \operatorname{Cov}\left[\mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi)\}}, \mathbb{1}_{\{1-U_i^{(n)} \geq \bar{\beta}_n(\xi)\}} \,\middle|\, \xi^{(n)}_{N,i} = \xi\right] - \eta_n^2 \; . \qquad \text{(A.13)}
\end{aligned}$$

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

$$\begin{aligned}
&\operatorname{Cov}\left[\mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi)\}}, \mathbb{1}_{\{1-U_i^{(n)} < \bar{\beta}_n(\xi)\}} \,\middle|\, \xi^{(n)}_{N,i} = \xi\right] \\
&\quad = \mathbb{P}\left(1 - \bar{\beta}_n(\xi) < U_i^{(n)} < \bar{\beta}_n(\xi) \,\middle|\, \xi^{(n)}_{N,i} = \xi\right) - \bar{\beta}_n^2(\xi) \qquad \text{(A.14)} \\
&\quad = \mathbb{1}_{\{\bar{\beta}_n(\xi) > 1/2\}}(2\bar{\beta}_n(\xi) - 1) - \bar{\beta}_n^2(\xi) \; ,
\end{aligned}$$

and, analogously,

$$\begin{aligned}
&\operatorname{Cov}\left[\mathbb{1}_{\{U_i^{(n)} < \bar{\beta}_n(\xi)\}}, \mathbb{1}_{\{1-U_i^{(n)} \geq \bar{\beta}_n(\xi)\}} \,\middle|\, \xi^{(n)}_{N,i} = \xi\right] \\
&\quad = \mathbb{P}\left(U_i^{(n)} \leq \min\{\bar{\beta}_n(\xi), 1 - \bar{\beta}_n(\xi)\} \,\middle|\, \xi^{(n)}_{N,i} = \xi\right) - \bar{\beta}_n(\xi)(1 - \bar{\beta}_n(\xi)) \qquad \text{(A.15)} \\
&\quad = \mathbb{1}_{\{\bar{\beta}_n(\xi) \leq 1/2\}}\bar{\beta}_n(\xi) + \mathbb{1}_{\{\bar{\beta}_n(\xi) > 1/2\}}(1 - \bar{\beta}_n(\xi)) - \bar{\beta}_n(\xi)(1 - \bar{\beta}_n(\xi)) \; .
\end{aligned}$$

Now, assume that $\bar{\beta}_n(\xi) > 1/2$; then, using (A.13)–(A.15),

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

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

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

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

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

the identity (4.10) follows.

References

Arvidsen, N. I., and Johnsson, T. (1982) Variance reduction through negative correlation—a simulation study. *J. Statist. Comput. Simulation* **15**, pp. 119–127.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994) ARCH Models. In *The Handbook of Econometrics* (eds. Engle, R. F., and McFadden, D.), **4**, pp. 2959–3038. Amsterdam: North-Holland.

Cappé, O., Moulines, É., and 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. *Ann. Statist.*, **32**, pp. 2385–2411.

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

Del Moral, P. (2004) *Feyman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications.* New York: Springer.

Douc, R., and Moulines, É. (2005) Limit theorems for weighted samples with applications to sequential Monte Carlo methods. To appear in *Ann. Stat.*

Douc, R., Moulines, É., and Olsson, J. (2008) Optimality of the auxiliary particle filter. To appear in *Prob. Math. Statist.*, **28**.

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

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

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

Kitagawa, G. (1987) Non-Gaussian state space modeling of nonstationary time series. *J. Am. Statist. Assoc.*, **82**, pp. 1023–1063.

Pitt, M. K., and Shephard, N. (1999) Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.*, **87**, pp. 493–499.

Polson, N. G., Stroud, J. R., and Müller, P. (2002) Practical filtering with sequential parameter learning. Technical report, University of Chicago.

(S. Bizjajeva) Lund University, Sweden
E-mail address: svetik@maths.lth.se

(J. Olsson) Lund University, Sweden
E-mail address: jimmy@maths.lth.se