PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 29, Fasc. 1 (2009), pp. 1–28

OPTIMALITY OF THE AUXILIARY PARTICLE FILTER

BY

RANDAL DOUC (PARIS), ÉRIC MOULINES (PARIS) AND JIMMY OLSSON (LUND)

Abstract. In this article we study asymptotic properties of weighted samples produced by the auxiliary particle filter (APF) proposed by Pitt and Shephard [17]. Besides establishing a central limit theorem (CLT) for smoothed particle estimates, we also derive bounds on the L^p error and bias of the same for a finite particle sample size. By examining the recursive formula for the asymptotic variance of the CLT we identify first-stage importance weights for which the increase of asymptotic variance at a single iteration of the algorithm is minimal. In the light of these findings, we discuss and demonstrate on several examples how the APF algorithm can be improved.

2000 AMS Mathematics Subject Classification: Primary: 65C05; Secondary: 65C60.

Key words and phrases: Auxiliary particle filter, central limit theorem, adjustment multiplier weight, sequential Monte Carlo, state space model, stratified sampling, two-stage sampling.

1. INTRODUCTION

In this paper we consider a *state space model* where a sequence $Y \triangleq \{Y_k\}_{k=0}^{\infty}$ is modeled as a noisy observation of a Markov chain $X \triangleq \{X_k\}_{k=0}^{\infty}$, called the *state sequence*, which is hidden. The observed values of Y are conditionally independent given the hidden states X and the corresponding conditional distribution of Y_k depends on X_k only. When operating on a model of this form the *joint smoothing distribution*, that is, the joint distribution of $(X_0, \ldots, X_n)$ given $(Y_0, \ldots, Y_n)$, and its marginals will be of interest. Of particular interest is the *filter distribution*, defined as the marginal of this law with respect to the component X_n is referred to. Computing these posterior distributions will be the key issue when filtering the hidden states as well as performing inference on unknown model parameters. The posterior distribution can be recursively updated as new observations become available – making single-sweep processing of the data possible – by means of the so-called *smoothing recursion*. However, in general, this recursion cannot be applied directly since it involves the evaluation of complicated high-dimensional integrals. In fact, closed form solutions are obtainable only for linear/Gaussian models (where the solutions are acquired using the *disturbance smoother*) and models where the state space of the latent Markov chain is finite.

Sequential Monte Carlo (SMC) *methods*, often alternatively termed *particle filters*, provide a helpful tool for computing approximate solutions to the smoothing recursion for general state space models, and the field has seen a drastic increase in interest over recent years. These methods are based on the principle of, recursively in time, approximating the smoothing distribution with the empirical measure associated with a weighted sample of *particles*. At present time there are various techniques for producing and updating such a particle sample (see [8], [6] and [13]). For a comprehensive treatment of the theoretical aspects of SMC methods we refer to the work by Del Moral [4].

In this article we analyse the *auxiliary particle filter* (APF) proposed by Pitt and Shephard [17], which has proved to be one of the most useful and widely adopted implementations of the SMC methodology. Unlike the traditional *bootstrap particle filter* [9], the APF enables the user to affect the particle sample allocation by designing freely a set of *first-stage importance weights* involved in the selection procedure. Prevalently, this has been used for assigning large weight to particles whose offsprings are likely to land up in zones of the state space having high posterior probability. Despite its obvious appeal, it is however not clear how to optimally exploit this additional degree of freedom.

In order to better understand this issue, we present an asymptotical analysis (being a continuation of [15] and based on recent results by [3], [12], [5] on weighted systems of particles) of the algorithm. More specifically, we establish CLTs (Theorems 3.1 and 3.2), with explicit expressions of the asymptotic variances, for two different versions (differentiated by the absence/presence of a concluding resampling pass at the end of each loop) of the algorithm under general model specifications. The convergence bear upon an increasing number of particles, and a recent result in the same spirit has, independently of [15], been stated in the manuscript [7]. Using these results, we also – and this is the main contribution of the paper – identify first-stage importance weights which are asymptotically most efficient. This result provides important insights in optimal sample allocation for particle filters in general, and we also give an interpretation of the finding in terms of variance reduction for stratified sampling.

In addition, we prove (utilising a decomposition of the Monte Carlo error proposed by Del Moral [4] and refined by Olsson et al. [14]) time uniform convergence in L^p (Theorem 3.3) under more stringent assumptions of ergodicity of the conditional hidden chain. With support of this stability result and the asymptotic analysis we conclude that inserting a final selection step at the end of each loop is – at least as long as the number of particles used in the two stages agree – superfluous, since such an operation exclusively increases the asymptotic variance.

Finally, in the implementation section (Section 5) several heuristics, derived from the obtained results, for designing efficient first-stage weights are discussed, and the improvement implied by approximating the asymptotically optimal first-stage weights is demonstrated on several examples.

2. NOTATION AND BASIC CONCEPTS

2.1. Model description. We denote by $(\mathsf{X}, \mathcal{X})$, Q, and ν the state space, transition kernel, and initial distribution of X, respectively, and assume that all random variables are defined on a common probability space $(\Omega, \mathbb{P}, \mathcal{A})$. In addition, we denote by $(\mathsf{Y}, \mathcal{Y})$ the state space of Y and suppose that there exists a measure λ and, for all $x \in \mathsf{X}$, a non-negative function $y \mapsto g(y|x)$ such that, for $k \geqslant 0$, $\mathbb{P}(Y_k \in A \mid X_k = x) = \int_A g(y|x)\lambda(dy)$, $A \in \mathcal{Y}$. Introduce, for $i \leqslant j$, the vector notation $\boldsymbol{X}_{i:j} \triangleq (X_i, \ldots, X_j)$; a similar notation will be used for other quantities. The joint smoothing distribution is denoted by

$$\phi_n(A) \triangleq \mathbb{P}(\boldsymbol{X}_{0:n} \in A \mid \boldsymbol{Y}_{0:n} = \boldsymbol{y}_{0:n}), \quad A \in \mathcal{X}^{\otimes(n+1)},$$

and a straightforward application of Bayes's formula shows that

$$\phi_{k+1}(A) = \frac{\int_A g(y_{k+1}|x_{k+1}) Q(x_k, dx_{k+1}) \phi_k(d\boldsymbol{x}_{0:k})}{\int_{\mathsf{X}^{k+2}} g(y_{k+1}|x'_{k+1}) Q(x'_k, dx'_{k+1}) \phi_k(d\boldsymbol{x}'_{0:k})} \tag{2.1}$$

for sets $A \in \mathcal{X}^{\otimes(k+2)}$. Throughout this paper we will assume that we are given a sequence $\{y_k; k \geqslant 0\}$ of *fixed* observations, and write, for $x \in \mathsf{X}$, $g_k(x) \triangleq g(y_k|x)$. Moreover, from now on we let the dependence on these observations of all other quantities be implicit, and denote, since the coming analysis is made exclusively *conditionally* on the given observed record, by $\mathbb{P}$ and $\mathbb{E}$ the conditional probability measure and expectation with respect to these observations.

2.2. The auxiliary particle filter. Let us recall the APF algorithm by Pitt and Shephard [17]. Assume that at time k we have a particle sample $\{(\boldsymbol{\xi}_{0:k}^{N,i}, \omega_k^{N,i})\}_{i=1}^N$ (each random variable $\boldsymbol{\xi}_{0:k}^{N,i}$ taking values in X^{k+1}) providing an approximation $\sum_{i=1}^N \omega_k^{N,i} \delta_{\boldsymbol{\xi}_{0:k}^{N,i}} / \Omega_k^N$ of the joint smoothing distribution ϕ_k with $\Omega_k^N \triangleq \sum_{i=1}^N \omega_k^{N,i}$ and $\omega_k^{N,i} \geqslant 0$, $1 \leqslant i \leqslant N$. Then, when the observation y_{k+1} becomes available, an approximation of ϕ_{k+1} is obtained by plugging this weighted empirical measure into the recursion (2.1), yielding

$$\bar{\phi}_{k+1}^N(A) \triangleq \sum_{i=1}^N \frac{\omega_k^{N,i} H_k^{\mathrm{u}}(\boldsymbol{\xi}_{0:k}^{N,i}, \mathsf{X}^{k+2})}{\sum_{j=1}^N \omega_k^{N,j} H_k^{\mathrm{u}}(\boldsymbol{\xi}_{0:k}^{N,j}, \mathsf{X}^{k+2})} H_k(\boldsymbol{\xi}_{0:k}^{N,i}, A), \quad A \in \mathcal{X}^{\otimes(k+2)}.$$

Here we have introduced, for $\boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}$ and $A \in \mathcal{X}^{\otimes(k+2)}$, the unnormalised kernels

$$H_k^{\mathrm{u}}(\boldsymbol{x}_{0:k}, A) \triangleq \int_A g_{k+1}(x'_{k+1}) \delta_{\boldsymbol{x}_{0:k}}(d\boldsymbol{x}'_{0:k}) Q(x'_k, dx'_{k+1})$$

and $H_k(\boldsymbol{x}_{0:k}, A) \triangleq H_k^{\mathrm{u}}(\boldsymbol{x}_{0:k}, A)/H_k^{\mathrm{u}}(\boldsymbol{x}_{0:k}, \mathsf{X}^{k+2})$. Simulating from $H_k(\boldsymbol{x}_{0:k}, A)$ consists in extending the trajectory $\boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}$ with an additional component being distributed according to the *optimal kernel*, that is, the distribution of X_{k+1} conditional on $X_k = x_k$ *and* the observation $Y_{k+1} = y_{k+1}$. Now, since we want to form a new weighted sample approximating ϕ_{k+1}, we need to find a convenient mechanism for sampling from $\bar{\phi}_{k+1}^N$ given $\{(\boldsymbol{\xi}_{0:k}^{N,i}, \omega_k^{N,i})\}_{i=1}^N$. In most cases it is possible – but generally computationally expensive – to simulate from $\bar{\phi}_{k+1}^N$ directly using *auxiliary accept-reject sampling* (see [11], [12]). A computationally cheaper solution (see [12], p. 1988, for a discussion of the acceptance probability associated with the auxiliary accept-reject sampling approach) consists in producing a weighted sample approximating $\bar{\phi}_{k+1}^N$ by sampling from the importance sampling distribution

$$\rho_{k+1}^N(A) \triangleq \sum_{i=1}^N \frac{\omega_k^{N,i}\tau_k^{N,i}}{\sum_{j=1}^N \omega_k^{N,j}\tau_k^{N,j}} R_k^{\mathrm{p}}(\boldsymbol{\xi}_{0:k}^{N,i}, A), \quad A \in \mathcal{X}^{\otimes(k+2)}.$$

Here $\tau_k^{N,i}$, $1 \leqslant i \leqslant N$, are positive numbers referred to as *first-stage weights* (Pitt and Shephard [17] use the term *adjustment multiplier weights*) and in this article we consider first-stage weights of type

$$\tau_k^{N,i} = t_k(\boldsymbol{\xi}_{0:k}^{N,i}) \tag{2.1}$$

for some function $t_k : \mathsf{X}^{k+1} \to \mathbb{R}^+$. Moreover, the pathwise proposal kernel R_k^{p} is, for $\boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}$ and $A \in \mathcal{X}^{\otimes(k+2)}$, of the form

$$R_k^{\mathrm{p}}(\boldsymbol{x}_{0:k}, A) = \int_A \delta_{\boldsymbol{x}_{0:k}}(d\boldsymbol{x}'_{0:k}) R_k(x'_k, dx'_{k+1})$$

with R_k being such that $Q(x, \cdot) \ll R_k(x, \cdot)$ for all $x \in \mathsf{X}$. Thus, a draw from $R_k^{\mathrm{p}}(\boldsymbol{x}_{0:k}, \cdot)$ is produced by extending the trajectory $\boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}$ with an additional component obtained by simulating from $R_k(x_k, \cdot)$. It is easily checked that for $\boldsymbol{x}_{0:k+1} \in \mathsf{X}^{k+2}$

$$\begin{aligned} \frac{d\bar{\phi}_{k+1}^N}{d\rho_{k+1}^N}(\boldsymbol{x}_{0:k+1}) &\propto w_{k+1}(\boldsymbol{x}_{0:k+1}) \\ &\triangleq \sum_{i=1}^N \mathbb{1}_{\boldsymbol{\xi}_{0:k}^{N,i}}(\boldsymbol{x}_{0:k}) \frac{g_{k+1}(x_{k+1})}{\tau_k^{N,i}} \frac{dQ(x_k, \cdot)}{dR_k(x_k, \cdot)}(x_{k+1}). \end{aligned} \tag{2.2}$$

An updated weighted particle sample $\{(\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}, \tilde{\omega}_{k+1}^{N,i})\}_{i=1}^{M_N}$ targeting $\bar{\phi}_{k+1}^N$ is hence generated by simulating M_N particles $\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}$, $1 \leqslant i \leqslant M_N$, from the proposal ρ_{k+1}^N and associating with these *second-stage weights* $\tilde{\omega}_{k+1}^{N,i} \triangleq w_{k+1}(\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i})$, $1 \leqslant$

$i \leqslant M_N$. By the identity function in (2.2), only a single term of the sum will contribute to the second-stage weight of a particle.

Finally, in an *optional* second-stage resampling pass a uniformly weighted particle sample $\{(\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}, 1)\}_{i=1}^{N}$, still targeting $\bar{\phi}_{k+1}^{N}$, is obtained by resampling N of the particles $\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}$, $1 \leqslant i \leqslant M_N$, according to the normalised second-stage weights. Note that the number of particles in the last two samples, M_N and N, may be different. The procedure is now repeated recursively (with $\omega_{k+1}^{N,i} \equiv 1$, $1 \leqslant i \leqslant N$) and is initialised by drawing $\boldsymbol{\xi}_0^{N,i}$, $1 \leqslant i \leqslant N$, independently of ς, where $\nu \ll \varsigma$, yielding $\omega_0^{N,i} = w_0(\boldsymbol{\xi}_0^{N,i})$ with $w_0(x) \triangleq g_0(x)\, d\nu/d\varsigma(x)$, $x \in \mathsf{X}$. To summarise, we obtain, depending on whether second-stage resampling is performed or not, the procedures described in Algorithms 1 and 2.

Algorithm 1 Two-Stage Sampling Particle Filter (TSSPF)

Ensure: $\{(\boldsymbol{\xi}_{0:k}^{N,i}, \omega_k^{N,i})\}_{i=1}^{N}$ approximates ϕ_k.

for $i = 1, \ldots, M_N$ **do** ▷ First stage

  draw indices $I_k^{N,i}$ from the set $\{1, \ldots, N\}$ multinomially with respect to the normalised weights $\omega_k^{N,j} \tau_k^{N,j} / \sum_{\ell=1}^{N} \omega_k^{N,\ell} \tau_k^{N,\ell}$, $1 \leqslant j \leqslant N$;

  simulate $\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}(k+1) \sim R_k[\boldsymbol{\xi}_{0:k}^{N,I_k^{N,i}}(k), \cdot]$, and

  set $\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i} \triangleq [\boldsymbol{\xi}_{0:k}^{N,I_k^{N,i}}, \tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}(k+1)]$ and $\tilde{\omega}_{k+1}^{N,i} \triangleq w_{k+1}(\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i})$.

end for

for $i = 1, \ldots, N$ **do** ▷ Second stage

  draw indices $J_{k+1}^{N,i}$ from the set $\{1, \ldots, M_N\}$ multinomially with respect to the normalised weights $\tilde{\omega}_{k+1}^{N,j} / \sum_{\ell=1}^{N} \tilde{\omega}_{k+1}^{N,\ell}$, $1 \leqslant j \leqslant N$, and

  set $\boldsymbol{\xi}_{0:k+1}^{N,i} \triangleq \tilde{\boldsymbol{\xi}}_{0:k+1}^{N,J_{k+1}^{N,i}}$.

  Finally, reset the weights: $\omega_{k+1}^{N,i} = 1$.

end for

Take $\{(\boldsymbol{\xi}_{0:k+1}^{N,i}, 1)\}_{i=1}^{N}$ as an approximation of ϕ_{k+1}.

We will use the term APF as a family name for both these algorithms and refer to them separately as *two-stage sampling particle filter* (TSSPF) and *single-stage auxiliary particle filter* (SSAPF). Note that by letting $\tau_k^{N,i} \equiv 1$, $1 \leqslant i \leqslant N$, in Algorithm 2 we obtain the bootstrap particle filter suggested by Gordon et al. [9].

The resampling steps of the APF can of course be implemented using techniques (e.g., *residual* or *systematic* resampling) different from multinomial resampling, leading to straightforward adaptations not discussed here. We believe however that the results of the coming analysis are generally applicable and extendable to a large class of selection schemes.

The issue whether second-stage resampling should be performed or not has been treated by several authors, and the theoretical results on the particle approxi-

Algorithm 2 Single-Stage Auxiliary Particle Filter (SSAPF)

Ensure: $\{(\boldsymbol{\xi}_{0:k}^{N,i}, \omega_k^{N,i})\}_{i=1}^N$ approximates ϕ_k.

for $i = 1, \ldots, N$ **do**
    draw indices $I_k^{N,i}$ from the set $\{1, \ldots, N\}$ multinomially with respect to the normalised weights $\omega_k^{N,j}\tau_k^{N,j} / \sum_{\ell=1}^N \omega_k^{N,\ell}\tau_k^{N,\ell}$, $1 \leqslant j \leqslant N$;
    simulate $\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}(k+1) \sim R_k[\boldsymbol{\xi}_{0:k}^{N,I_k^{N,i}}(k), \cdot]$, and
    set $\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i} \triangleq [\boldsymbol{\xi}_{0:k}^{N,I_k^{N,i}}, \tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}(k+1)]$ and $\tilde{\omega}_{k+1}^{N,i} \triangleq w_{k+1}(\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i})$.
end for
Take $\{(\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}, \tilde{\omega}_{k+1}^{N,i})\}_{i=1}^N$ as an approximation of ϕ_{k+1}.

mation stability and asymptotic variance presented in the next section will indicate that the second-stage selection pass should, at least for the case $M_N = N$, be canceled, since this exclusively increases the sampling variance. Thus, the idea that the second-stage resampling pass is necessary for preventing the particle approximation from degenerating does not apparently hold. Recently, a similar conclusion was reached in the manuscript [7].

The advantage of the APF not possessed by standard SMC methods is the possibility of, firstly, choosing the first-stage weights $\tau_k^{N,i}$ arbitrarily and, secondly, letting N and M_N be different (TSSPF only). Appealing to common sense, SMC methods work efficiently when the particle weights are well-balanced, and Pitt and Shephard [17] propose several strategies for achieving this by adapting the first-stage weights. In some cases it is possible to fully adapt the filter to the model (see Section 5), providing exactly equal importance weights; otherwise, Pitt and Shephard [17] suggest, in the case $R_k \equiv Q$ and $\mathsf{X} = \mathbb{R}^d$, the generic first-stage importance weight function

$$t_k^{\mathrm{P\&S}}(\boldsymbol{x}_{0:k}) \triangleq g_{k+1}\Big[\int_{\mathbb{R}^d} x' Q(x_k, dx')\Big], \qquad \boldsymbol{x}_{0:k} \in \mathbb{R}^{k+1}.$$

The analysis that follows will however show that this way of adapting the first-stage weights is not necessarily good in terms of asymptotic (as N tends to infinity) sample variance; indeed, using first-stage weights given by $t_k^{\mathrm{P\&S}}$ can be even detrimental for some models.

3. BOUNDS AND ASYMPTOTICS FOR PRODUCED APPROXIMATIONS

3.1. Asymptotic properties. Introduce, for any probability measure μ on some measurable space $(\mathsf{E}, \mathcal{E})$ and μ-measurable function f satisfying $\int_{\mathsf{E}} |f(x)|\mu(dx) < \infty$, the notation $\mu f \triangleq \int_{\mathsf{E}} f(x)\mu(dx)$. Moreover, for any two transition kernels K and T from $(\mathsf{E}_1, \mathcal{E}_1)$ to $(\mathsf{E}_2, \mathcal{E}_2)$ and $(\mathsf{E}_2, \mathcal{E}_2)$ to $(\mathsf{E}_3, \mathcal{E}_3)$, respectively, we define the product transition kernel $KT(x, A) \triangleq \int_{\mathsf{E}_2} T(z, A)K(x, dz)$ for $x \in \mathsf{E}_1$ and $A \in \mathcal{E}_3$. A set C of real-valued functions on X^m 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 $|f| \leqslant |g|$, then $|f| \in \mathsf{C}$; (iii) for all $c \in \mathbb{R}$, the constant function $f \equiv c$ belongs to C.

From [5] we adapt the following definitions.

DEFINITION 3.1 (Consistency). A weighted sample $\{(\boldsymbol{\xi}_{0:m}^{N,i}, \omega_m^{N,i})\}_{i=1}^{M_N}$ on the space X^{m+1} is said to be *consistent* for the probability measure μ and the (proper) set $\mathsf{C} \subseteq \mathsf{L}^1(\mathsf{X}^{m+1}, \mu)$ if, for any $f \in \mathsf{C}$, as $N \to \infty$,

$$(\Omega_m^N)^{-1} \sum_{i=1}^{M_N} \omega_m^{N,i} f(\boldsymbol{\xi}_{0:m}^{N,i}) \xrightarrow{\mathbb{P}} \mu f,$$
$$(\Omega_m^N)^{-1} \max_{1 \leqslant i \leqslant M_N} \omega_m^{N,i} \xrightarrow{\mathbb{P}} 0.$$

DEFINITION 3.2 (Asymptotic normality). A sample $\{(\boldsymbol{\xi}_{0:m}^{N,i}, \omega_m^{N,i})\}_{i=1}^{M_N}$ on X^{m+1} is called *asymptotically normal* for $(\mu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{a_N\}_{N=1}^{\infty})$ if, as $N \to \infty$,

$$a_N (\Omega_m^N)^{-1} \sum_{i=1}^{M_N} \omega_m^{N,i} [f(\boldsymbol{\xi}_{0:m}^{N,i}) - \mu f] \xrightarrow{\mathcal{D}} \mathcal{N}[0, \sigma^2(f)] \quad \text{for any } f \in \mathsf{A},$$
$$a_N^2 (\Omega_m^N)^{-1} \sum_{i=1}^{M_N} (\omega_m^{N,i})^2 f(\boldsymbol{\xi}_{0:m}^{N,i}) \xrightarrow{\mathbb{P}} \gamma f \quad \text{for any } f \in \mathsf{W},$$
$$a_N (\Omega_m^N)^{-1} \max_{1 \leqslant i \leqslant M_N} \omega_m^{N,i} \xrightarrow{\mathbb{P}} 0.$$

The main contribution of this section are the following results, which establish consistency and asymptotic normality of weighted samples produced by the TSSPF and SSAPF algorithms. For all $k \geqslant 0$, we define a transformation Φ_k on the set of ϕ_k-integrable functions by

$$\Phi_k[f](\boldsymbol{x}_{0:k}) \triangleq f(\boldsymbol{x}_{0:k}) - \phi_k f, \quad \boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}. \tag{3.1}$$

In addition, we impose the following assumptions:

(A1) *For all* $k \geqslant 1$, $t_k \in \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k)$ *and* $w_k \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k)$, *where* t_k *and* w_k *are defined in* (2.1) *and* (2.2), *respectively.*

(A2) (i) $\mathsf{A}_0 \subseteq \mathsf{L}^1(\mathsf{X}, \phi_0)$ *is a proper set and* $\sigma_0 : \mathsf{A}_0 \to \mathbb{R}^+$ *is a function satisfying, for all* $f \in \mathsf{A}_0$ *and* $a \in \mathbb{R}$,

$$\sigma_0(af) = |a|\sigma_0(f).$$

(ii) *The initial sample* $\{(\boldsymbol{\xi}_0^{N,i}, 1)\}_{i=1}^{N}$ *is consistent for* $[\mathsf{L}^1(\mathsf{X}, \phi_0), \phi_0]$ *and asymptotically normal for* $[\phi_0, \mathsf{A}_0, \mathsf{W}_0, \sigma_0, \gamma_0, \{\sqrt{N}\}_{N=1}^{\infty}]$.

THEOREM 3.1. *Assume* (A1) *and* (A2) *with* $(\mathsf{W}_0, \gamma_0) = [\mathsf{L}^1(\mathsf{X}, \phi_0), \phi_0]$. *In the setting of Algorithm* 1, *suppose that the limit* $\beta \triangleq \lim_{N\to\infty} N/M_N$ *exists, where* $\beta \in [0,1]$. *Define recursively the family* $\{\mathsf{A}_k\}_{k=1}^{\infty}$ *by*

$$\text{(3.2)}\quad \mathsf{A}_{k+1} \triangleq \{f \in \mathsf{L}^2(\mathsf{X}^{k+2}, \phi_{k+1}) : R_k^{\mathrm{p}}(\cdot, w_{k+1}|f|) H_k^{\mathrm{u}}(\cdot, |f|) \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k),$$
$$H_k^{\mathrm{u}}(\cdot, |f|) \in \mathsf{A}_k \cap \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k), w_{k+1} f^2 \in \mathsf{L}^1(\mathsf{X}^{k+2}, \ \phi_{k+1})\}.$$

Moreover, define recursively the family $\{\sigma_k\}_{k=1}^{\infty}$ *of functionals* $\sigma_k : \mathsf{A}_k \to \mathbb{R}^+$ *by*

$$\text{(3.3)}\quad \sigma_{k+1}^2(f) \triangleq \phi_{k+1}\Phi_{k+1}^2[f] + \frac{\sigma_k^2\{H_k^{\mathrm{u}}(\cdot, \Phi_{k+1}[f])\} + \beta\phi_k\{t_k R_k^{\mathrm{p}}(\cdot, w_{k+1}^2 \Phi_{k+1}^2[f])\}\, \phi_k t_k}{[\phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})]^2}.$$

Then all sets A_k, $k \geqslant 1$, *are proper; moreover, all samples* $\{(\boldsymbol{\xi}_{0:k}^{N,i}, 1)\}_{i=1}^{N}$ *produced by Algorithm* 1 *are consistent and asymptotically normal for* $[\mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k), \phi_k]$ *and* $[\phi_k, \mathsf{A}_k, \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k), \sigma_k, \phi_k, \{\sqrt{N}\}_{N=1}^{\infty}]$, *respectively.*

The proof is given in Section 6, and as a by-product a similar result for the SSAPF (Algorithm 2) is obtained.

THEOREM 3.2. *Assume* (A1) *and* (A2). *Define the families* $\{\tilde{\mathsf{W}}_k\}_{k=0}^{\infty}$ *and* $\{\tilde{\mathsf{A}}_k\}_{k=0}^{\infty}$ *by*

$$\tilde{\mathsf{W}}_k \triangleq \{f \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k) : w_{k+1} f \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k)\}, \qquad \tilde{\mathsf{W}}_0 \triangleq \mathsf{W}_0,$$

and, with $\tilde{\mathsf{A}}_0 \triangleq \mathsf{A}_0$,

$$\text{(3.4)}\quad \tilde{\mathsf{A}}_{k+1} \triangleq \{f \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}) : R_k^{\mathrm{p}}(\cdot, w_{k+1}|f|) H_k^{\mathrm{u}}(\cdot, |f|) \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k),$$
$$H_k^{\mathrm{u}}(\cdot, |f|) \in \tilde{\mathsf{A}}_k, [H_k^{\mathrm{u}}(\cdot, |f|)]^2 \in \tilde{\mathsf{W}}_k, w_{k+1} f^2 \in \mathsf{L}^1(\mathsf{X}^{k+2}, \ \phi_{k+1})\}.$$

Moreover, define recursively the family $\{\tilde{\sigma}_k\}_{k=0}^{\infty}$ *of functionals* $\tilde{\sigma}_k : \mathsf{A}_k \to \mathbb{R}^+$ *by*

$$\text{(3.5)}\quad \tilde{\sigma}_{k+1}^2(f) \triangleq \frac{\tilde{\sigma}_k^2\{H_k^{\mathrm{u}}(\cdot, \Phi_{k+1}[f])\} + \phi_k\{t_k R_k^{\mathrm{p}}(\cdot, w_{k+1}^2 \Phi_{k+1}^2[f])\}\, \phi_k t_k}{[\phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})]^2}, \quad \tilde{\sigma}_0 \triangleq \sigma_0,$$

and the measures $\{\tilde{\gamma}_k\}_{k=1}^{\infty}$ *by*

$$\tilde{\gamma}_{k+1} f \triangleq \frac{\phi_{k+1}(w_{k+1} f)\, \phi_k t_k}{\phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})}, \qquad f \in \tilde{\mathsf{W}}_{k+1}.$$

Then all $\tilde{\mathsf{A}}_k$, $k \geqslant 1$, *are proper; moreover, all samples* $\{(\tilde{\boldsymbol{\xi}}_{0:k}^{N,i}, \tilde{\omega}_k^{N,i})\}_{i=1}^{N}$ *produced by Algorithm* 2 *are consistent and asymptotically normal for* $[\mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k), \phi_k]$ *and* $[\phi_k, \tilde{\mathsf{A}}_k, \tilde{\mathsf{W}}_k, \tilde{\sigma}_k, \tilde{\gamma}_k, \{\sqrt{N}\}_{N=1}^{\infty}]$, *respectively.*

Under the assumption of bounded likelihood and second-stage importance weight functions g_k and w_k, one can show that the CLTs stated in Theorems 3.1 and 3.2 indeed include any functions having finite second moments with respect to the joint smoothing distributions; that is, under these assumptions the supplementary constraints on the sets (3.2) and (3.4) are automatically fulfilled. This is the contents of the statement below.

(A3) *For all* $k \geqslant 0$, $\|g_k\|_{\mathsf{X},\infty} < \infty$ *and* $\|w_k\|_{\mathsf{X}^{k+1},\infty} < \infty$.

COROLLARY 3.1. *Assume* (A3) *and let* $\{\mathsf{A}_k\}_{k=0}^{\infty}$ *and* $\{\tilde{\mathsf{A}}_k\}_{k=0}^{\infty}$ *be defined by* (3.2) *and* (3.4), *respectively*, *with* $\tilde{\mathsf{A}}_0 = \mathsf{A}_0 \triangleq \mathsf{L}^2(\mathsf{X}, \phi_0)$. *Then*, *for all* $k \geqslant 1$, $\mathsf{A}_k = \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k)$ *and* $\mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k) \subseteq \tilde{\mathsf{A}}_k$.

For a proof, see Section 6.2.

Interestingly, the expressions of $\tilde{\sigma}^2_{k+1}(f)$ and $\sigma^2_{k+1}(f)$ differ, for $\beta = 1$, *only on the additive term* $\phi_{k+1}\Phi^2_{k+1}[f]$, that is, the variance of f under ϕ_{k+1}. This quantity represents the cost of introducing the second-stage resampling pass, which was proposed as a mean for preventing the particle approximation from degenerating. In the coming Section 3.2 we will however show that the approximations produced by the SSAPF are already stable for a finite time horizon, and that additional resampling is superfluous. Thus, there are indeed reasons for strongly questioning whether second-stage resampling should be performed at all, at least when the same number of particles are used in the two stages.

3.2. Bounds on L^p error and bias. In this part we examine, under suitable regularity conditions and for a finite particle population, the errors of the approximations obtained by the APF in terms of L^p bounds and bounds on the bias. We preface our main result with some definitions and assumptions. Denote by $\mathcal{B}_{\mathrm{b}}(\mathsf{X}^m)$ a space of bounded measurable functions on X^m furnished with the supremum norm $\|f\|_{\mathsf{X}^m,\infty} \triangleq \sup_{\boldsymbol{x}\in\mathsf{X}^m} |f(\boldsymbol{x})|$. Let, for $f \in \mathcal{B}_{\mathrm{b}}(\mathsf{X}^m)$, the *oscillation seminorm* (alternatively termed the *global modus of continuity*) be defined by $\operatorname{osc}(f) \triangleq \sup_{(\boldsymbol{x},\boldsymbol{x}')\in\mathsf{X}^m\times\mathsf{X}^m} |f(\boldsymbol{x}) - f(\boldsymbol{x}')|$. Furthermore, the L^p norm of a stochastic variable X is denoted by $\|X\|_p \triangleq \mathbb{E}^{1/p}[|X|^p]$. When considering sums, we will make use of the standard convention $\sum_{k=a}^{b} c_k = 0$ if $b < a$.

In the following we will assume that all measures $Q(x, \cdot)$, $x \in \mathsf{X}$, have densities $q(x, \cdot)$ with respect to a common dominating measure μ on $(\mathsf{X}, \mathcal{X})$. Moreover, we suppose that the following holds.

(A4) (i) $\epsilon_- \triangleq \inf_{(x,x')\in\mathsf{X}^2} q(x, x') > 0$, $\epsilon_+ \triangleq \sup_{(x,x')\in\mathsf{X}^2} q(x, x') < \infty$.

(ii) *For all* $y \in \mathsf{Y}$, $\int_{\mathsf{X}} g(y|x)\,\mu(dx) > 0$.

Under (A4) we define

$$\rho \triangleq 1 - \frac{\epsilon_-}{\epsilon_+}\,. \tag{3.6}$$

(A5) *For all* $k \geqslant 0$, $\|t_k\|_{\mathsf{X}^{k+1},\infty} < \infty$.

Assumption (A4) is now standard and is often satisfied when the state space X is compact and implies that the hidden chain, when evolving conditionally on the observations, is geometrical ergodic with a mixing rate given by $\rho < 1$. For comprehensive treatments of such stability properties within the framework of state space models we refer to Del Moral [4]. Finally, let $\mathcal{C}_i(\mathsf{X}^{n+1})$ be the set of bounded measurable functions f on X^{n+1} of type $f(\boldsymbol{x}_{0:n}) = \bar{f}(\boldsymbol{x}_{i:n})$ for some function $\bar{f} : \mathsf{X}^{n-i+1} \to \mathbb{R}$. In this setting we have the following result, which is proved in Section 6.3.

THEOREM 3.3. *Assume* (A3), (A4), (A5), *and let* $f \in \mathcal{C}_i(\mathsf{X}^{n+1})$ *for* $0 \leqslant i \leqslant n$. *Let* $\{(\tilde{\boldsymbol{\xi}}_{0:k}^{N,i}, \tilde{\omega}_k^{N,i})\}_{i=1}^{R_N(r)}$ *be a weighted particle sample produced by Algorithm* r, $r = \{1, 2\}$, *with* $R_N(r) \triangleq \mathbb{1}\{r = 1\}M_N + \mathbb{1}\{r = 2\}N$. *Then the following holds true for all* $N \geqslant 1$ *and* $r = \{1, 2\}$.

(i) *For all* $p \geqslant 2$,

$$\Big\|(\tilde{\Omega}_n^N)^{-1} \sum_{j=1}^{R_N(r)} \tilde{\omega}_n^{N,j} f_i(\tilde{\boldsymbol{\xi}}_{0:n}^{N,j}) - \phi_n f_i\Big\|_p$$

$$\leqslant B_p \frac{\operatorname{osc}(f_i)}{1-\rho} \Bigg[\frac{1}{\epsilon_- \sqrt{R_N(r)}} \sum_{k=1}^{n} \frac{\|w_k\|_{\mathsf{X}^{k+1},\infty} \|t_{k-1}\|_{\mathsf{X}^k,\infty}}{\mu g_k} \rho^{0 \vee (i-k)} + \frac{\mathbb{1}\{r=1\}}{\sqrt{N}} \left(\frac{\rho}{1-\rho} + n - i \right) + \frac{\|w_0\|_{\mathsf{X},\infty}}{\nu g_0 \sqrt{N}} \rho^i \Bigg].$$

(ii) *We have*

$$\Big|\mathbb{E}\big[(\tilde{\Omega}_n^N)^{-1} \sum_{j=1}^{R_N(r)} \tilde{\omega}_n^{N,j} f_i(\tilde{\boldsymbol{\xi}}_{0:n}^{N,j})\big] - \phi_n f_i\Big|$$

$$\leqslant B \frac{\operatorname{osc}(f_i)}{(1-\rho)^2} \Bigg[\frac{1}{R_N(r)\epsilon_-^2} \sum_{k=1}^{n} \frac{\|w_k\|_{\mathsf{X}^{k+1},\infty}^2 \|t_{k-1}\|_{\mathsf{X}^k,\infty}^2}{(\mu g_k)^2} \rho^{0 \vee (i-k)} + \frac{\mathbb{1}\{r=1\}}{N} \left(\frac{\rho}{1-\rho} + n - i \right) + \frac{\|w_0\|_{\mathsf{X},\infty}^2}{N(\nu g_0)^2} \rho^i \Bigg].$$

Here ρ *is defined in* (3.6), *and* B_p *and* B *are universal constants such that* B_p *depends on* p *only.*

Especially, assuming that all fractions $\|w_k\|_{\mathsf{X}^{k+1},\infty} \|t_{k-1}\|_{\mathsf{X}^k,\infty} / \mu g_k$ are uniformly bounded in k and applying Theorem 3.3 for $i = n$ yields error bounds on the approximate filter distribution which are *uniformly bounded* in n. From this it is obvious that the first-stage resampling pass is enough to preserve the sample stability. Indeed, by avoiding second-stage selection according to Algorithm 2 we can obtain, since the middle terms in the bounds above cancel in this case, even *tighter* control of the L^p error for a fixed number of particles.

4. IDENTIFYING ASYMPTOTICALLY OPTIMAL FIRST-STAGE WEIGHTS

The formulas (3.3) and (3.5) for the asymptotic variances of the TSSPF and SSAPF may look complicated at a first sight, but by careful examining the same we will obtain important knowledge of how to choose the first-stage importance weight functions t_k in order to robustify the APF .

Assume that we have run the APF up to time k and are about to design suitable first-stage weights for the next iteration. In this setting, we call a first-stage weight function $t'_k[f]$, possibly depending on the target function $f \in \mathsf{A}_{k+1}$ and satisfying (A1), *optimal* (at time k) if it provides a minimal increase of asymptotic variance at a single iteration of the APF algorithm, that is, if $\sigma^2_{k+1}\{t'_k[f]\}(f) \leqslant \sigma^2_{k+1}\{t\}(f)$ (or $\tilde{\sigma}^2_{k+1}\{t'_k[f]\}(f) \leqslant \tilde{\sigma}^2_{k+1}\{t\}(f)$) for all other measurable and positive weight functions t. Here we let $\sigma^2_{k+1}\{t\}(f)$ denote the asymptotic variance induced by t. Define, for $\boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}$,

$$
\text{(4.1)} \quad t^*_k[f](\boldsymbol{x}_{0:k}) \triangleq \sqrt{\int_{\mathsf{X}} g^2_{k+1}(x_{k+1}) \left[\frac{dQ(x_k, \cdot)}{dR_k(x_k, \cdot)}(x_{k+1}) \right]^2 \Phi^2_{k+1}[f](\boldsymbol{x}_{0:k+1}) R_k(x_k, dx_{k+1})},
$$

and let $w^*_{k+1}[f]$ denote the second-stage importance weight function induced by $t^*_k[f]$ according to (2.2). We are now ready to state the main result of this section. The proof is found in Section 6.4.

THEOREM 4.1. *Let $k \geqslant 0$ and define t^*_k by* (4.1). *Then the following is valid*:

(i) *Let the assumptions of Theorem* 3.1 *hold and suppose that $f \in \{f' \in \mathsf{A}_{k+1} \colon t^*_k[f'] \in \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k), w^*_{k+1}[f'] \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1})\}$. Then t^*_k is optimal for Algorithm* 1 *and the corresponding minimal variance is given by*

$$
\sigma^2_{k+1}\{t^*_k\}(f) = \phi_{k+1}\Phi^2_{k+1}[f] + \frac{\sigma^2_k\big[H^{\mathrm{u}}_k(\cdot, \Phi_{k+1}[f])\big] + \beta(\phi_k t^*_k[f])^2}{[\phi_k H^{\mathrm{u}}_k(\mathsf{X}^{k+2})]^2}.
$$

(ii) *Let the assumptions of Theorem* 3.2 *hold and suppose that $f \in \{f' \in \tilde{\mathsf{A}}_{k+1} : t^*_k[f'] \in \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k), w^*_{k+1}[f'] \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1})\}$. Then t^*_k is optimal for Algorithm* 2 *and the corresponding minimal variance is given by*

$$
\tilde{\sigma}^2_{k+1}\{t^*_k\}(f) = \frac{\tilde{\sigma}^2_k\big[H^{\mathrm{u}}_k(\cdot, \Phi_{k+1}[f])\big] + (\phi_k t^*_k[f])^2}{[\phi_k H^{\mathrm{u}}_k(\mathsf{X}^{k+2})]^2}.
$$

The functions t^*_k have a natural interpretation in terms of optimal sample allocation for *stratified sampling*. Consider the mixture $\pi = \sum_{i=1}^d w_i \mu_i$, each μ_i being a measure on some measurable space $(\mathsf{E}, \mathcal{E})$ and $\sum_{i=1}^d w_i = 1$, and the problem of estimating, for some given π-integrable target function f, the expectation πf. In order to relate this to the particle filtering paradigm, we will make use of Algorithm 3.

Algorithm 3 Stratified importance sampling

for $i = 1, \ldots, N$ **do**
  draw an index J_i multinomially with respect to τ_j, $1 \leqslant j \leqslant d$, so that $\sum_{j=1}^{d} \tau_j = 1$;
  simulate $\xi_i \sim \nu_{J_i}$, and
  compute the weights $\omega_i \triangleq \left. \frac{w_j}{\tau_j} \frac{d\mu_j}{d\nu_j} \right|_{j=J_i}$
end for
Take $\{(\xi_i, \omega_i)\}_{i=1}^{N}$ as an approximation of π.

In other words, we perform Monte Carlo estimation of πf by means of sampling from some proposal mixture $\sum_{j=1}^{d} \tau_j \nu_j$ and forming a self-normalised estimate; cf. the technique applied in Section 2.2 for sampling from $\bar{\phi}_{k+1}^{N}$. In this setting, the following CLT can be established under weak assumptions:

$$\sqrt{N}\left[\sum_{i=1}^{N} \frac{\omega_i}{\sum_{\ell=1}^{N} \omega_\ell} f(\xi_i) - \pi f\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left[0, \sum_{j=1}^{d} \frac{w_j^2 \alpha_j(f)}{\tau_j}\right]$$

with, for $x \in \mathsf{E}$,

$$\alpha_i(f) \triangleq \int_{\mathsf{E}} \left[\frac{d\mu_i}{d\nu_i}(x)\right]^2 \Pi^2[f](x)\, \nu_i(dx) \quad \text{and} \quad \Pi[f](x) \triangleq f(x) - \pi f.$$

Minimising the asymptotic variance $\sum_{i=1}^{d} [w_i^2 \alpha_i(f)/\tau_i]$ with respect to τ_i, $1 \leqslant i \leqslant d$, e.g., by means of the Lagrange multiplicator method (the details are simple), yields the optimal weights

$$\tau_i^* \propto w_i \sqrt{\alpha_i(f)} = w_i \sqrt{\int_{\mathsf{E}} \left[\frac{d\mu_i}{d\nu_i}(x)\right]^2 \Pi^2[f](x)\, \nu_i(dx)},$$

and the similarity between this expression and that of the optimal first-stage importance weight functions t_k^* is striking. This strongly supports the idea of interpreting optimal sample allocation for particle filters in terms of variance reduction for stratified sampling.

5. IMPLEMENTATIONS

As shown in the previous section, the utilisation of the optimal weights (4.1) provides, for a given sequence $\{R_k\}_{k=0}^{\infty}$ of proposal kernels, the most efficient of all particle filters belonging to the large class covered by Algorithm 2 (including the standard bootstrap filter and any fully adapted particle filter). However, exact computation of the optimal weights is in general infeasible by two reasons: firstly, they depend (via $\Phi_{k+1}[f]$) on the expectation $\phi_{k+1} f$, that is, the quantity that we aim to estimate, and, secondly, they involve the evaluation of a complicated integral. A comprehensive treatment of the important issue of how to approximate the optimal weights is beyond the scope of this paper, but in the following three examples we discuss some possible heuristics for doing this.

5.1. Nonlinear Gaussian model. In order to form an initial idea of the performance of the optimal SSAPF in practice, we apply the method to a first order (possibly nonlinear) autoregressive model observed in noise:

$$
\begin{aligned}
X_{k+1} &= m(X_k) + \sigma_w(X_k) W_{k+1}, \\
Y_k &= X_k + \sigma_v V_k,
\end{aligned}
\tag{5.1}
$$

with $\{W_k\}_{k=1}^{\infty}$ and $\{V_k\}_{k=0}^{\infty}$ being mutually independent sets of standard normal distributed variables such that W_{k+1} is independent of (X_i, Y_i), $0 \leqslant i \leqslant k$, and V_k is independent of X_k, (X_i, Y_i), $0 \leqslant i \leqslant k-1$. Here the functions $\sigma_w : \mathbb{R} \to \mathbb{R}^+$ and $m : \mathbb{R} \to \mathbb{R}$ are measurable, and $\mathsf{X} = \mathbb{R}$. As observed by Pitt and Shephard [17], it is, for all models of the form (5.1), possible to propose a new particle using the optimal kernel directly, yielding $R_k^{\mathrm{p}} = H_k$ and, for $(x, x') \in \mathbb{R}^2$,

$$
r_k(x, x') = \frac{1}{\tilde{\sigma}_k(x)\sqrt{2\pi}} \exp\left\{ -\frac{[x' - \tilde{m}_k(x)]^2}{2\tilde{\sigma}_k^2(x)} \right\},
\tag{5.2}
$$

with r_k denoting the density of R_k with respect to the Lebesgue measure, and

$$
\tilde{m}_k(x) \triangleq \left[\frac{y_{k+1}}{\sigma_v^2} + \frac{m_k(x)}{\sigma_w^2(x)} \right] \tilde{\sigma}_k^2(x), \qquad \tilde{\sigma}_k^2(x) \triangleq \frac{\sigma_v^2 \sigma_w^2(x)}{\sigma_v^2 + \sigma_w^2(x)}.
\tag{5.3}
$$

For the proposal (5.2) it is, for $\boldsymbol{x}_{k:k+1} \in \mathbb{R}^2$, valid that

$$
\begin{aligned}
g_{k+1}(x_{k+1}) \frac{dQ(x_k, \cdot)}{dR_k(x_k, \cdot)}(x_{k+1}) &\propto h_k(x_k) \\
&\triangleq \frac{\tilde{\sigma}_k(x_k)}{\sigma_w(x_k)} \exp\left[\frac{\tilde{m}_k^2(x_k)}{2\tilde{\sigma}_k^2(x_k)} - \frac{m^2(x_k)}{2\sigma_w^2(x_k)} \right],
\end{aligned}
\tag{5.4}
$$

and since the right-hand side does not depend on x_{k+1}, we can obtain, by letting $t_k(\boldsymbol{x}_{0:k}) = h_k(x_k)$, $\boldsymbol{x}_{0:k} \in \mathbb{R}^{k+1}$, second-stage weights being indeed unity (providing a sample of genuinely $\bar{\phi}_{k+1}^N$-distributed particles). When this is achieved, Pitt and Shephard [17] call the particle filter *fully adapted*. There is however nothing in the previous theoretical analysis that supports the idea that aiming at evenly distributed second-stage weights is always convenient, and this will also be illustrated in the simulations below. On the other hand, it is possible to find *cases* when the fully adapted particle filter is very close to being optimal; see again the following discussion.

In the following subsections we will study two special cases of (5.1).

5.2. Linear/Gaussian model. Consider the case

$$m(X_k) = \phi X_k \quad \text{and} \quad \sigma_w(X_k) \equiv \sigma.$$

For a linear/Gaussian model of this kind, exact expressions of the optimal weights can be obtained using the Kalman filter. We set $\phi = 0.9$ and let the latent chain be put at stationarity from the beginning, that is, $X_0 \sim \mathcal{N}[0, \sigma^2/(1 - \phi^2)]$. In this setting, we simulated, for $\sigma = \sigma_v = 0.1$, a record $\boldsymbol{y}_{0:10}$ of observations and estimated the filter posterior means (corresponding to projection target functions $\pi_k(\boldsymbol{x}_{0:k}) \triangleq x_k$, $\boldsymbol{x}_{0:k} \in \mathbb{R}^{k+1}$) along this trajectory by applying (1) SSAPF based on true optimal weights, (2) SSAPF based on the generic weights $t_k^{\mathrm{P\&S}}$ of Pitt and Shephard [17], and (3) the standard bootstrap particle filter (that is, SSAPF with $t_k \equiv 1$). In this first experiment, the prior kernel Q was taken as proposal in all cases, and since the optimal weights are derived using asymptotic arguments, we used as many as $100{,}000$ particles for all algorithms. The result is displayed in Figure 1 (a), and it is clear that operating with true optimal allocation weights improves – as expected – the MSE performance in comparison with the other methods.

The main motivation of Pitt and Shephard [17] for introducing auxiliary particle filtering was to robustify the particle approximation to outliers. Thus, we mimic Cappé et al. [2], Example 7.2.3, and repeat the experiment above for the observation record $\boldsymbol{y}_{0:5} = (-0.652, -0.345, -0.676, 1.142, 0.721, 20)$, standard deviations $\sigma_v = 1$, $\sigma = 0.1$, and the smaller particle sample size $N = 10{,}000$. Note the large discrepancy of the last observation y_5, which in this case is located at a distance of 20 standard deviations from the mean of the stationary distribution. The outcome is plotted in Figure 1 (b) from which it is evident that the particle filter based on the optimal weights is the most efficient also in this case; moreover, the performance of the standard auxiliary particle filter is improved in comparison with the bootstrap filter. Figure 2 displays a plot of the weight functions t_4^* and $t_4^{\mathrm{P\&S}}$ for the same observation record. It is clear that $t_4^{\mathrm{P\&S}}$ is not too far away from the optimal weight function (which is close to symmetric in this extreme situation) in this case, even if the distance between the functions as measured with the supremum norm is still significant.

Finally, we implement the fully adapted filter (with proposal kernels and first-stage weights given by (5.2) and (5.4), respectively) and compare this with the SSAPF based on the same proposal (5.4) and optimal first-stage weights, the latter being given, for $\boldsymbol{x}_{0:k} \in \mathbb{R}^{k+1}$ and h_k defined in (5.4), by

$$\begin{aligned}(5.5) \quad t_k^*[\pi_{k+1}](\boldsymbol{x}_{0:k}) &\propto h_k(x_k) \sqrt{\int_{\mathbb{R}} \Phi_{k+1}^2[\pi_{k+1}](\boldsymbol{x}_{k+1}) R_k(x_k, dx_{k+1})} \\ &= h_k(x_k) \sqrt{\tilde{\sigma}_k^2(x_k) + \tilde{m}_k^2(x_k) - 2\tilde{m}_k(x_k)\phi_{k+1}\pi_{k+1} + \phi_{k+1}^2 \pi_{k+1}}\end{aligned}$$

in this case. We note that h_k, that is, the first-stage weight function for the fully adapted filter, enters as a factor in the optimal weight function (5.5). Moreover,

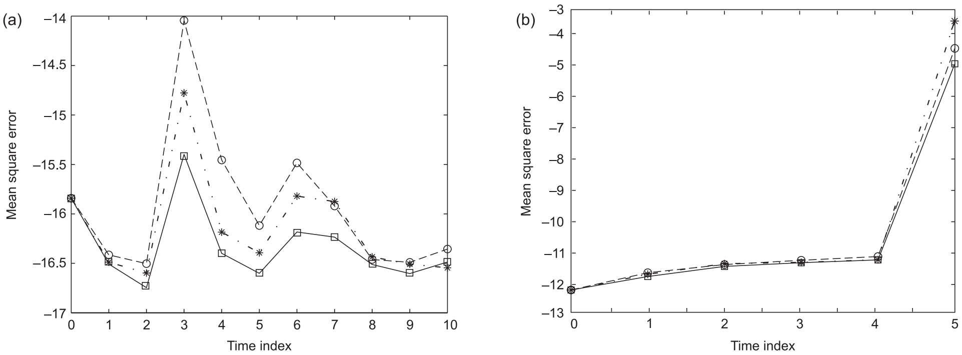

FIGURE 1. Plot of MSE perfomances (on log-scale) of the bootstrap particle filter ($*$), the SSAPF based on optimal weights ($\Box$), and the SSAPF based on the generic weights $t_k^{\mathrm{P\&S}}$ of Pitt and Shephard [17] ($\circ$). The MSE values are founded on 100,000 particles and 400 runs of each algorithm

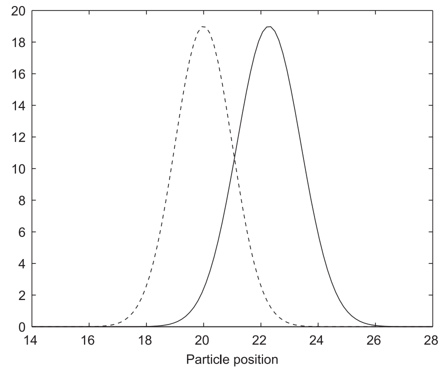

FIGURE 2. Plot of the first-stage importance weight functions t_4^* (unbroken line) and $t_4^{\mathrm{P\&S}}$ (dashed line) in the presence of an outlier

recall the definitions (5.3) of $\tilde{m}_k$ and $\tilde{\sigma}_k$; in the case of very informative observations, corresponding to $\sigma_v \ll \sigma$, it holds that $\tilde{\sigma}_k(x) \approx \sigma_v$ and $\tilde{m}_k(x) \approx y_{k+1}$ with good precision for moderate values of $x \in \mathbb{R}$ (that is, values not too far away from the mean of the stationary distribution of X). Thus, the factor beside h_k in (5.5) is more or less constant in this case, implying that the fully adapted and optimal first-stage weight filters are close to equivalent. This observation is perfectly confirmed in Figure 3 (a) which presents MSE performances for $\sigma_v = 0.1$, $\sigma = 1$, and $N = 10{,}000$. In the same figure, the bootstrap filter and the standard auxiliary filter based on generic weights are included for a comparison, and these (particularly the latter) are marred with significantly larger Monte Carlo errors. On the contrary, in the case of non-informative observations, that is, $\sigma_v \gg \sigma$, we note that $\tilde{\sigma}_k(x) \approx \sigma$, $\tilde{m}_k(x) \approx \phi x$ and conclude that the optimal kernel is close the prior kernel Q. In addition, the exponent of h_k vanishes, implying uniform first-stage weights for the fully adapted particle filter. Thus, the fully adapted filter will be close to the bootstrap filter in this case, and Figure 3 (b) seems to confirm this remark. Moreover, the optimal first-stage weight filter does clearly better than the others in terms of MSE performance.

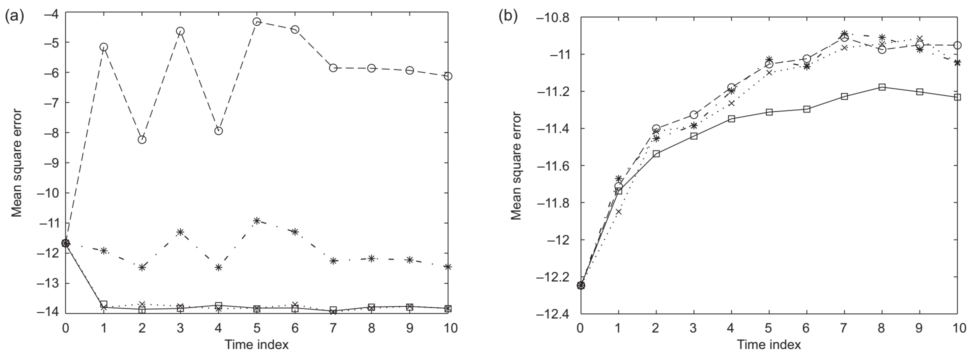

FIGURE 3. Plot of MSE perfomances (on log-scale) of the bootstrap particle filter ($*$), the SSAPF based on optimal weights ($\square$), the SSAPF based on the generic weights $t_k^{\mathrm{P\&S}}$ ($\circ$), and the fully adapted SSAPF ($\times$) for the linear/Gaussian model in Section 5.2. The MSE values are computed using 10,000 particles and 400 runs of each algorithm

5.3. ARCH model. Now, let instead

$$m(X_k) \equiv 0 \quad \text{and} \quad \sigma_w(X_k) = \sqrt{\beta_0 + \beta_1 X_k^2}.$$

Here we deal with the classical Gaussian autoregressive conditional heteroscedasticity (ARCH) model (see [1]) observed in noise. Since the nonlinear state equation precludes exact computation of the filtered means, implementing the optimal first-stage weight SSAPF is considerably more challenging in this case. The problem can however be tackled by means of an introductory *zero-stage* simulation pass, based on $R \ll N$ particles, in which a crude estimate of $\phi_{k+1} f$ is obtained. For instance, this can be achieved by applying the standard bootstrap filter with multinomial resampling. Using this approach, we computed again MSE values for the bootstrap filter, the standard SSAPF based on generic weights, the fully adapted SSAPF, and the (approximate) optimal first-stage weight SSAPF, the latter using the optimal proposal kernel. Each algorithm used $10{,}000$ particles and the number of particles in the prefatory pass was set to $R = N/10 = 1000$, implying only a minor additional computational work. An imitation of the true filter means was obtained by running the bootstrap filter with as many as $500{,}000$ particles. In compliance with the foregoing, we considered the case of informative (Figure 4 (a)) as well as non-informative (Figure 4 (b)) observations, corresponding to $(\beta_0, \beta_1, \sigma_v) = (9, 5, 1)$ and $(\beta_0, \beta_1, \sigma_v) = (0.1, 1, 3)$, respectively. Since $\tilde{\sigma}_k(x) \approx \sigma_v$, $\tilde{m}_k(x) \approx y_{k+1}$ in the latter case, we should, in accordance with the previous discussion, again expect the fully adapted filter to be close to that based on optimal first-stage weights. This is also confirmed in the plot. For the former parameter set, the fully adapted SSAPF exhibits an MSE performance close to that of the bootstrap filter, while the optimal first-stage weight SSAPF is clearly superior.

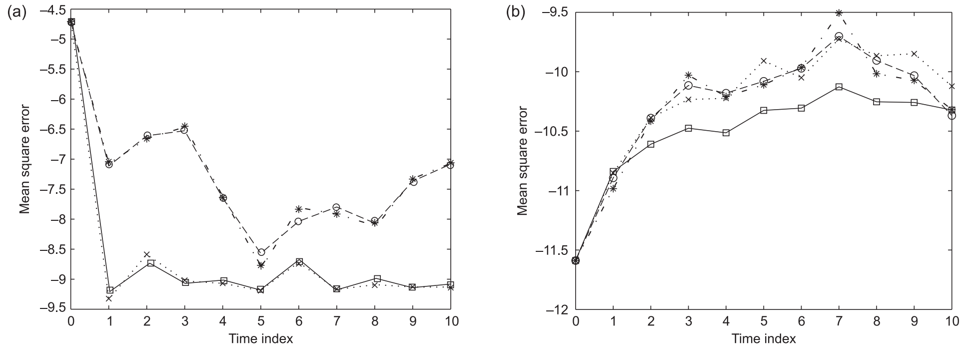

FIGURE 4. Plot of MSE perfomances (on log-scale) of the bootstrap particle filter ($*$), the SSAPF based on optimal weights ($\square$), the SSAPF based on the generic weights $t_k^{\mathrm{P\&S}}$ ($\circ$), and the fully adapted SSAPF ($\times$) for the ARCH model in Section 5.3. The MSE values are computed using 10,000 particles and 400 runs of each algorithm

5.4. Stochastic volatility. As a final example let us consider the canonical discrete-time *stochastic volatility* (SV) *model* [10] given by

$$\begin{aligned} X_{k+1} &= \phi X_k + \sigma W_{k+1}, \\ Y_k &= \beta \exp(X_k/2) V_k, \end{aligned}$$

where $\mathsf{X} = \mathbb{R}$, and $\{W_k\}_{k=1}^{\infty}$ and $\{V_k\}_{k=0}^{\infty}$ are as in Example 5.1. Here X and Y are log-volatility and log-returns, respectively, where the former are assumed to be stationary. Also this model was treated by Pitt and Shephard [17], who discussed approximate full adaptation of the particle filter by means of a second order Taylor approximation of the concave function $x' \mapsto \log g_{k+1}(x')$. More specifically, by multiplying the approximate observation density obtained in this way with $q(x, x')$, $(x, x') \in \mathbb{R}^2$, yielding a Gaussian approximation of the optimal kernel density, nearly even second-stage weights can be obtained. We proceed in the same spirit, approximating however directly the (log-concave) function $x' \mapsto g_{k+1}(x') q(x, x')$ by means of a second order Taylor expansion of $x' \mapsto \log[g_{k+1}(x') q(x, x')]$ around the mode $\bar{m}_k(x)$ (obtained using Newton iterations) of the same:

$$\begin{aligned} & g_{k+1}(x') q(x, x') \\ & \approx r_k^{\mathrm{u}}(x, x') \triangleq g_{k+1}[\bar{m}_k(x)] q[x, \bar{m}_k(x)] \exp\left\{ -\frac{1}{2\bar{\sigma}_k^2(x)} [x' - \bar{m}_k(x)]^2 \right\}, \end{aligned}$$

with (we refer to [2], pp. 225–228, for details) $\bar{\sigma}_k^2(x)$ being the inverted negative of the second order derivative, evaluated at $\bar{m}_k(x)$, of $x' \mapsto \log[g_{k+1}(x') q(x, x')]$.

Thus, by letting, for $(x, x') \in \mathbb{R}^2$, $r_k(x, x') = r_k^{\mathrm{u}}(x, x') / \int_{\mathbb{R}} r_k^{\mathrm{u}}(x, x'') dx''$, we obtain

$$(5.6) \quad g_{k+1}(x_{k+1}) \frac{dQ(x_k, \cdot)}{dR_k(x_k, \cdot)}(x_{k+1}) \approx \int_{\mathbb{R}} r_k^{\mathrm{u}}(x_k, x')\, dx' \propto \bar{\sigma}_k(x_k) g_{k+1}[\bar{m}_k(x_k)] q[x_k, \bar{m}_k(x_k)],$$

and letting, for $\boldsymbol{x}_{0:k} \in \mathbb{R}^{k+1}$, $t_k(\boldsymbol{x}_{0:k}) = \bar{\sigma}_k(x_k) g_{k+1}[\bar{m}_k(x_k)] q[x_k, \bar{m}_k(x_k)]$ will imply a nearly fully adapted particle filter. Moreover, by applying the approximate relation (5.6) to the expression (4.1) of the optimal weights, we get (cf. (5.5))

$$(5.7) \quad t_k^*[\pi_{k+1}](\boldsymbol{x}_{0:k}) \approx \int_{\mathbb{R}} r_k^{\mathrm{u}}(x_k, x') dx' \sqrt{\int_{\mathbb{R}} \Phi_{k+1}^2[\pi_{k+1}](x) R_k(x_k, dx)} \propto \sqrt{\bar{\sigma}_k^2(x_k) + \bar{m}_k^2(x_k) - 2\bar{m}_k(x_k)\phi_{k+1}\pi_{k+1} + \phi_{k+1}^2 \pi_{k+1}} \times \bar{\sigma}_k(x_k) g_{k+1}[\bar{m}_k(x_k)] q[x_k, \bar{m}_k(x_k)].$$

In this setting, a numerical experiment was conducted where the two filters above were run, again together with the bootstrap filter and the auxiliary filter based on the generic weights $t_k^{\mathrm{P\&S}}$, for parameters $(\phi, \beta, \sigma) = (0.9702, 0.5992, 0.178)$ (estimated by Pitt and Shephard [18] from daily returns on the U.S. dollar against the U.K. pound stearling from the first day of trading in 1997 and for the next 200 days). To make the filtering problem more challenging, we used a simulated record $\boldsymbol{y}_{0:10}$ of observations arising from the initial state $x_0 = 2.19$, being above the 2% quantile of the stationary distribution of X, implying a sequence of relatively impetuously fluctuating log-returns. The number of particles was set to $N = 5000$ for all filters, and the number of particles used in the prefatory filtering

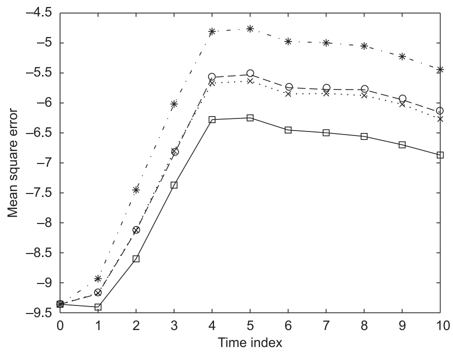

FIGURE 5. Plot of MSE perfomances (on log-scale) of the bootstrap particle filter ($*$), the SSAPF based on optimal weights ($\square$), the SSAPF based on the generic weights $t_k^{\mathrm{P\&S}}$ ($\circ$), and the fully adapted SSAPF ($\times$) for the SV model in Section 5.4. The MSE values are computed using 5000 particles and 400 runs of each algorithm

pass (in which a rough approximation of $\phi_{k+1}\pi_{k+1}$ in (5.7) was computed using the bootstrap filter) of the SSAPF filter based on optimal first-stage weights was set to $R = N/5 = 1000$; thus, running the optimal first-stage weight filter is only marginally more demanding than running the fully adapted filter. The outcome is displayed in Figure 5. It is once more obvious that introducing approximate optimal first-stage weights significantly improves the performance also for the SV model, which is recognised as being specially demanding as regards state estimation.

6. APPENDIX — PROOFS

6.1. Proof of Theorem 3.1. Let us recall the updating scheme described in Algorithm 1 and formulate it in the following four isolated steps:

$$
\begin{aligned}
(6.1)\quad & \{(\boldsymbol{\xi}_{0:k}^{N,i}, 1)\}_{i=1}^{N} \xrightarrow{\text{I: Weighting}} \{(\boldsymbol{\xi}_{0:k}^{N,i}, \tau_k^{N,i})\}_{i=1}^{N} \to \\
& \xrightarrow{\text{II: Resampling (1st stage)}} \{(\hat{\boldsymbol{\xi}}_{0:k}^{N,i}, 1)\}_{i=1}^{M_N} \xrightarrow{\text{III: Mutation}} \{(\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}, \tilde{\omega}_{k+1}^{N,i})\}_{i=1}^{M_N} \to \\
& \xrightarrow{\text{IV: Resampling (2nd stage)}} \{(\boldsymbol{\xi}_{0:k+1}^{N,i}, 1)\}_{i=1}^{N},
\end{aligned}
$$

where we have set $\hat{\boldsymbol{\xi}}_{0:k}^{N,i} \triangleq \boldsymbol{\xi}_{0:k}^{N,I_k^{N,i}}$, $1 \leqslant i \leqslant M_N$. Now, the asymptotic properties stated in Theorem 3.1 are established by a chain of applications of Theorems 1–4 in [5]. We will proceed by induction: assume that the uniformly weighted particle sample $\{(\boldsymbol{\xi}_{0:k}^{N,i}, 1)\}_{i=1}^{N}$ is consistent for $[\mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k), \phi_k]$ and asymptotically normal for $[\phi_k, \mathsf{A}_k, \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k), \sigma_k, \phi_k, \{\sqrt{N}\}_{N=1}^{\infty}]$, with A_k being a proper set and σ_k such that $\sigma_k(af) = |a|\sigma_k(f)$, $f \in \mathsf{A}_k$, $a \in \mathbb{R}$. We prove, by analysing each of the steps I–IV, that this property is preserved through one iteration of the algorithm.

I. Define the measure

$$\mu_k(A) \triangleq \frac{\phi_k(t_k \mathbb{1}_A)}{\phi_k t_k}, \quad A \in \mathcal{X}^{\otimes(k+1)}.$$

Using Theorem 1 of [5] for $R(\boldsymbol{x}_{0:k}, \cdot) = \delta_{\boldsymbol{x}_{0:k}}(\cdot)$, $L(\boldsymbol{x}_{0:k}, \cdot) = t_k(\boldsymbol{x}_{0:k})\,\delta_{\boldsymbol{x}_{0:k}}(\cdot)$, $\mu = \mu_k$, and $\nu = \phi_k$, we conclude that the sample $\{(\boldsymbol{\xi}_{0:k}^{N,i}, \tau_k^{N,i})\}_{i=1}^{N}$ is consistent for $[\{f \in \mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k) : t_k|f| \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k)\}, \mu_k] = [\mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k), \mu_k]$. Here the equality is based on the fact that $\phi_k(t_k|f|) = \mu_k|f|\,\phi_k t_k$, where the second factor on the right-hand side is bounded by Assumption (A1). In addition, by applying Theorem 1 of [5] we conclude that $\{(\boldsymbol{\xi}_{0:k}^{N,i}, \tau_k^{N,i})\}_{i=1}^{N}$ is asymptotically normal for $(\mu_k, \mathsf{A}_{\mathrm{I},k}, \mathsf{W}_{\mathrm{I},k}, \sigma_{\mathrm{I},k}, \gamma_{\mathrm{I},k}, \{\sqrt{N}\}_{N=1}^{\infty})$, where

$$
\begin{aligned}
\mathsf{A}_{\mathrm{I},k} &\triangleq \{f \in \mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k) : t_k|f| \in \mathsf{A}_k, t_k f \in \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k)\} \\
&= \{f \in \mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k) : t_k f \in \mathsf{A}_k \cap \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k)\}, \\
\mathsf{W}_{\mathrm{I},k} &\triangleq \{f \in \mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k) : t_k^2|f| \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k)\}
\end{aligned}
$$

are proper sets, and

$$\sigma_{\mathrm{I},k}^2(f) \triangleq \sigma_k^2\left[\frac{t_k(f-\mu_k f)}{\phi_k t_k}\right] = \frac{\sigma_k^2[t_k(f-\mu_k f)]}{(\phi_k t_k)^2}, \quad f \in \mathsf{A}_{\mathrm{I},k},$$

$$\gamma_{\mathrm{I},k} f \triangleq \frac{\phi_k(t_k^2 f)}{(\phi_k t_k)^2}, \quad f \in \mathsf{W}_{\mathrm{I},k}.$$

II. By Theorems 3 and 4 of [5], $\{(\hat{\boldsymbol{\xi}}_{0:k}^{N,i}, 1)\}_{i=1}^{M_N}$ is consistent and asymptotically normal for $[\mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k), \mu_k]$ and $[\mu_k, \mathsf{A}_{\mathrm{II},k}, \mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k), \sigma_{\mathrm{II},k}, \beta\mu_k, \{\sqrt{N}\}_{N=1}^\infty]$, respectively, where

$$\begin{aligned}\mathsf{A}_{\mathrm{II},k} &\triangleq \{f \in \mathsf{A}_{\mathrm{I},k} : f \in \mathsf{L}^2(\mathsf{X}^{k+1}, \mu_k)\}\\ &= \{f \in \mathsf{L}^2(\mathsf{X}^{k+1}, \mu_k) : t_k f \in \mathsf{A}_k \cap \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k)\}\end{aligned}$$

is a proper set, and

$$\begin{aligned}\sigma_{\mathrm{II},k}^2(f) &\triangleq \beta\mu_k[(f-\mu_k f)^2] + \sigma_{\mathrm{I},k}^2(f)\\ &= \beta\mu_k[(f-\mu_k f)^2] + \frac{\sigma_k^2[t_k(f-\mu_k f)]}{(\phi_k t_k)^2}, \quad f \in \mathsf{A}_{\mathrm{II},k}.\end{aligned}$$

III. We argue as in step I, but this time for $\nu = \mu_k$, $R = R_k^{\mathrm{p}}$, and $L(\cdot, A) = R_k^{\mathrm{p}}(\cdot, w_{k+1}\mathbb{1}_A)$, $A \in \mathcal{X}^{\otimes(k+2)}$, providing the target distribution

$$\mu(A) = \frac{\mu_k R_k^{\mathrm{p}}(w_{k+1}\mathbb{1}_A)}{\mu_k R_k^{\mathrm{p}} w_{k+1}} = \frac{\phi_k H_k^{\mathrm{u}}(A)}{\phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})} = \phi_{k+1}(A), \quad A \in \mathcal{X}^{\otimes(k+2)}. \tag{6.2}$$

This yields, applying Theorems 1 and 2 of [5], that $\{(\tilde{\boldsymbol{\xi}}_{k+1}^{N,i}, \tilde{\omega}_{k+1}^{N,i})\}_{i=1}^{M_N}$ is consistent for

$$\begin{aligned}&[\{f \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}), R_k^{\mathrm{p}}(\cdot, w_{k+1}|f|) \in \mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k)\}, \phi_{k+1}]\\ &\qquad = [\mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}), \phi_{k+1}],\end{aligned} \tag{6.3}$$

where (6.3) follows, since $\mu_k R_k^{\mathrm{p}}(w_{k+1}|f|)\,\phi_k t_k = \phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})\,\phi_{k+1}|f|$, from (A1), and asymptotically normal for $(\phi_{k+1}, \mathsf{A}_{\mathrm{III},k+1}, \mathsf{W}_{\mathrm{III},k+1}, \sigma_{\mathrm{III},k+1}, \gamma_{\mathrm{III},k+1}, \{\sqrt{N}\}_{N=1}^\infty)$. Here

$$\begin{aligned}\mathsf{A}_{\mathrm{III},k+1} &\triangleq \{f \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}) : R_k^{\mathrm{p}}(\cdot, w_{k+1}|f|) \in \mathsf{A}_{\mathrm{II},k},\\ &\qquad\qquad R_k^{\mathrm{p}}(\cdot, w_{k+1}^2 f^2) \in \mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k)\}\\ &= \{f \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}) : R_k^{\mathrm{p}}(\cdot, w_{k+1}|f|) \in \mathsf{L}^2(\mathsf{X}^{k+1}, \mu_k),\\ &\quad t_k R_k^{\mathrm{p}}(\cdot, w_{k+1}|f|) \in \mathsf{A}_k \cap \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k), R_k^{\mathrm{p}}(\cdot, w_{k+1}^2 f^2) \in \mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k)\}\\ &= \{f \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}) : R_k^{\mathrm{p}}(\cdot, w_{k+1}|f|) H_k^{\mathrm{u}}(\cdot, |f|) \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k),\\ &\qquad H_k^{\mathrm{u}}(\cdot, |f|) \in \mathsf{A}_k \cap \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k), w_{k+1} f^2 \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1})\}\end{aligned}$$

and

$$\begin{aligned}\mathsf{W}_{\mathrm{III},k+1} &\triangleq \{f \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}) : R_k^{\mathrm{p}}(\cdot, w_{k+1}^2|f|) \in \mathsf{L}^1(\mathsf{X}^{k+1}, \mu_k)\} \\ &= \{f \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}) : w_{k+1} f \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1})\}\end{aligned}$$

are proper sets. In addition, from the identity (6.2) we obtain

$$\mu_k R_k^{\mathrm{p}}(w_{k+1}\Phi_{k+1}[f]) = 0,$$

where Φ_{k+1} is defined in (3.1), yielding, for $f \in \mathsf{A}_{\mathrm{III},k+1}$,

$$\begin{aligned}\sigma^2_{\mathrm{III},k+1}(f) &\triangleq \sigma^2_{\mathrm{II},k}\left\{\frac{R_k^{\mathrm{p}}(\cdot, w_{k+1}\Phi_{k+1}[f])}{\mu_k R_k^{\mathrm{p}} w_{k+1}}\right\} \\ &\qquad + \frac{\beta\mu_k R_k^{\mathrm{p}}\big(\{w_{k+1}\Phi_{k+1}[f] - R_k^{\mathrm{p}}(\cdot, w_{k+1}\Phi_{k+1}[f])\}^2\big)}{(\mu_k R_k^{\mathrm{p}} w_{k+1})^2} \\ &= \frac{\beta\mu_k\big(\{R_k^{\mathrm{p}}(w_{k+1}\Phi_{k+1}[f])\}^2\big)}{(\mu_k R_k^{\mathrm{p}} w_{k+1})^2} + \frac{\sigma_k^2\{t_k R_k^{\mathrm{p}}(\cdot, w_{k+1}\Phi_{k+1}[f])\}}{(\phi_k t_k)^2(\mu_k R_k^{\mathrm{p}} w_{k+1})^2} \\ &\qquad + \frac{\beta\mu_k R_k^{\mathrm{p}}\big(\{w_{k+1}\Phi_{k+1}[f] - R_k^{\mathrm{p}}(\cdot, w_{k+1}\Phi_{k+1}[f])\}^2\big)}{(\mu_k R_k^{\mathrm{p}} w_{k+1})^2}.\end{aligned}$$

Now, applying the equality

$$\begin{aligned}\{R_k^{\mathrm{p}}(\cdot, w_{k+1}\Phi_{k+1}[f])\}^2 + R_k^{\mathrm{p}}\big(\cdot, \{w_{k+1}\Phi_{k+1}[f] - R_k^{\mathrm{p}}(\cdot, w_{k+1}\Phi_{k+1}[f])\}^2\big) \\ = R_k^{\mathrm{p}}(\cdot, w_{k+1}^2\Phi_{k+1}^2[f])\end{aligned}$$

provides, for $f \in \mathsf{A}_{\mathrm{III},k+1}$, the variance

$$\sigma^2_{\mathrm{III},k+1}(f) = \frac{\beta\phi_k\{t_k R_k^{\mathrm{p}}(\cdot, w_{k+1}^2\Phi_{k+1}^2[f])\}\,\phi_k t_k + \sigma_k^2\{H_k^{\mathrm{u}}(\cdot, \Phi_{k+1}[f])\}}{[\phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})]^2}. \tag{6.4}$$

Finally, for $f \in \mathsf{W}_{\mathrm{III},k+1}$,

$$\gamma_{\mathrm{III},k+1} f \triangleq \frac{\beta\mu_k R_k^{\mathrm{p}}(w_{k+1}^2 f)}{(\mu_k R_k^{\mathrm{p}} w_{k+1})^2} = \frac{\beta\phi_{k+1}(w_{k+1}f)\,\phi_k t_k}{\phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})}.$$

IV. The consistency for $[\mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}), \phi_{k+1}]$ of the uniformly weighted particle sample $\{(\boldsymbol{\xi}_{0:k+1}^{N,i}, 1)\}_{i=1}^N$ follows from Theorem 3 in [5]. In addition, applying Theorem 4 of [5] yields that the same sample is asymptotically normal for $[\phi_{k+1}, \mathsf{A}_{\mathrm{IV},k+1}, \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1}), \sigma_{\mathrm{IV},k+1}, \phi_{k+1}, \{\sqrt{N}\}_{N=1}^\infty]$, with

$$\begin{aligned}\mathsf{A}_{\mathrm{IV},k+1} &\triangleq \{f \in \mathsf{A}_{\mathrm{III},k+1} : f \in \mathsf{L}^2(\mathsf{X}^{k+2}, \phi_{k+1})\} \\ &= \{f \in \mathsf{L}^2(\mathsf{X}^{k+2}, \phi_{k+1}) : R_k^{\mathrm{p}}(\cdot, w_{k+1}|f|) H_k^{\mathrm{u}}(\cdot, |f|) \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k), \\ &\qquad H_k^{\mathrm{u}}(\cdot, |f|) \in \mathsf{A}_k \cap \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k), w_{k+1} f^2 \in \mathsf{L}^1(\mathsf{X}^{k+2}, \phi_{k+1})\}\end{aligned}$$

being a proper set, and, for $f \in \mathsf{A}_{\mathrm{IV},k+1}$,

$$\sigma^2_{\mathrm{IV},k+1}(f) \triangleq \phi_{k+1}\Phi^2_{k+1}[f] + \sigma^2_{\mathrm{III},k+1}(f),$$

with $\sigma^2_{\mathrm{III},k+1}(f)$ being defined by (6.4). This concludes the proof of the theorem.

6.2. Proof of Corollary 3.1. We pick $f \in \mathsf{L}^2(\mathsf{X}^{k+2}, \phi_{k+1})$ and prove that the constraints of the set A_{k+1} defined in (3.2) are satisfied under Assumption (A3). Firstly, by Jensen's inequality,

$$\begin{aligned}\phi_k[R^{\mathrm{p}}_k(\cdot, w_{k+1}|f|)H^{\mathrm{u}}_k(\cdot, |f|)] &= \phi_k\{t_k[R^{\mathrm{p}}_k(\cdot, w_{k+1}|f|)]^2\} \\ &\leqslant \phi_k[t_k R^{\mathrm{p}}_k(\cdot, w^2_{k+1}f^2)] = \phi_k H^{\mathrm{u}}_k(w_{k+1}f^2) \\ &\leqslant \|w_{k+1}\|_{\mathsf{X}^{k+2},\infty}\, \phi_k H^{\mathrm{u}}_k(\mathsf{X}^{k+2})\, \phi_{k+1}(f^2) < \infty,\end{aligned}$$

and, similarly,

$$\phi_k\{[H^{\mathrm{u}}_k(\cdot, |f|)]^2\} \leqslant \|g_{k+1}\|_{\mathsf{X},\infty}\, \phi_k H^{\mathrm{u}}_k(\mathsf{X}^{k+2})\, \phi_{k+1}(f^2) < \infty.$$

From this, together with the bound

$$\phi_{k+1}(w_{k+1}f^2) \leqslant \|w_{k+1}\|_{\mathsf{X}^{k+2},\infty}\, \phi_{k+1}(f^2) < \infty,$$

we conclude that $\mathsf{A}_{k+1} = \mathsf{L}^2(\mathsf{X}^{k+2}, \phi_{k+1})$.

To prove $\mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k) \subseteq \tilde{\mathsf{A}}_k$, note that Assumption (A3) implies the equality $\tilde{\mathsf{W}}_k = \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k)$ and repeat the arguments above.

6.3. Proof of Theorem 3.3. Define, for $r \in \{1, 2\}$ and $R_N(r)$ as determined in Theorem 3.3, the empirical measures

$$\phi^N_k(A) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{\xi}^{N,i}_{0:k}}, \quad \tilde{\phi}^N_k(A) \triangleq \sum_{i=1}^{R_N(r)} \frac{\tilde{\omega}^{N,i}_k}{\tilde{\Omega}^N_k}\delta_{\tilde{\boldsymbol{\xi}}^{N,i}_{0:k}}(A), \quad A \in \mathcal{X}^{\otimes(k+1)},$$

playing the role of approximations of the smoothing distribution ϕ_k. Let us define $\mathcal{F}_0 \triangleq \sigma(\boldsymbol{\xi}^{N,i}_0; 1 \leqslant i \leqslant N)$; then the particle history up to the different steps of loop $m+1$, $m \geqslant 0$, of Algorithm r, $r \in \{1, 2\}$, is modeled by the filtrations $\hat{\mathcal{F}}_m \triangleq \mathcal{F}_m \vee \sigma[I^{N,i}_m; 1 \leqslant i \leqslant R_N(r)]$, $\tilde{\mathcal{F}}_{m+1} \triangleq \mathcal{F}_m \vee \sigma[\tilde{\boldsymbol{\xi}}^{N,i}_{0:m+1}; 1 \leqslant i \leqslant R_N(r)]$, and

$$\mathcal{F}_{m+1} \triangleq \begin{cases}\tilde{\mathcal{F}}_{m+1} \vee \sigma(J^{N,i}_{m+1}; 1 \leqslant i \leqslant N) & \text{for } r = 1,\\ \tilde{\mathcal{F}}_{m+1} & \text{for } r = 2,\end{cases}$$

respectively. In the coming proof we will describe one iteration of the APF algorithm by the following two operations:

$$\begin{aligned}&\{(\boldsymbol{\xi}^{N,i}_{0:k}, \omega^{N,i}_k)\}^N_{i=1} \xrightarrow{\text{Sampling from } \varphi^N_{k+1}} \{(\tilde{\boldsymbol{\xi}}^{N,i}_{0:k+1}, \tilde{\omega}^{N,i}_{k+1})\}^{R_N(r)}_{i=1} \rightarrow \\ &\qquad \xrightarrow{r\,=\,1\text{: Sampling from } \tilde{\phi}^N_{0:k+1}} \{(\boldsymbol{\xi}^{N,i}_{0:k+1}, 1)\}^N_{i=1},\end{aligned}$$

where, for $A \in \mathcal{X}^{\otimes(k+2)}$,

$$
\begin{aligned}
(6.5) \qquad \varphi_{k+1}^N(A) &\triangleq \mathbb{P}(\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i_0} \in A | \mathcal{F}_k) \\
&= \sum_{j=1}^N \frac{\omega_k^{N,j} \tau_k^{N,j}}{\sum_{\ell=1}^N \omega_k^{N,\ell} \tau_k^{N,\ell}} R_k^{\mathrm{p}}(\boldsymbol{\xi}_{0:k}^{N,j}, A) = \frac{\phi_k^N[t_k R_k^{\mathrm{p}}(\cdot, A)]}{\phi_k^N t_k}
\end{aligned}
$$

for some index $i_0 \in \{1, \ldots, R_N(r)\}$ (given $\mathcal{F}_k$, the particles $\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,i}$, $1 \leqslant i \leqslant R_N(r)$, are i.i.d.). Here the initial weights $\{\omega_k^{N,i}\}_{i=1}^N$ are all equal to one for $r = 1$. The second operation is valid since, for any $i_0 \in \{1, \ldots, N\}$,

$$
\mathbb{P}(\boldsymbol{\xi}_{0:k+1}^{N,i_0} \in A | \tilde{\mathcal{F}}_{k+1}) = \sum_{j=1}^{R_N(r)} \frac{\tilde{\omega}_{k+1}^{N,j}}{\tilde{\Omega}_{k+1}^N} \delta_{\tilde{\boldsymbol{\xi}}_{0:k+1}^{N,j}}(A) = \tilde{\phi}_{0:k+1}^N(A), \quad A \in \mathcal{X}^{\otimes(k+2)}.
$$

The fact that the evolution of the particles can be described by two Monte Carlo operations involving conditionally i.i.d. variables makes it possible to analyse the error using the Marcinkiewicz–Zygmund inequality (see [16], p. 62).

Using this, set, for $1 \leqslant k \leqslant n$,

$$
(6.6) \qquad \alpha_k^N(A) \triangleq \int_A \frac{d\alpha_k^N}{d\varphi_k^N}(\boldsymbol{x}_{0:k}) \varphi_k^N(d\boldsymbol{x}_{0:k}), \quad A \in \mathcal{X}^{\otimes(k+1)},
$$

with, for $\boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}$,

$$
\frac{d\alpha_k^N}{d\varphi_k^N}(\boldsymbol{x}_{0:k}) \triangleq \frac{w_k(\boldsymbol{x}_{0:k}) H_k^{\mathrm{u}} \ldots H_{n-1}^{\mathrm{u}}(\boldsymbol{x}_{0:k}, \mathsf{X}^{n+1})\, \phi_{k-1}^N t_{k-1}}{\phi_{k-1}^N H_{k-1}^{\mathrm{u}} \ldots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})}.
$$

Here we apply the standard convention $H_\ell^{\mathrm{u}} \ldots H_m^{\mathrm{u}} \triangleq \mathrm{Id}$ if $m < \ell$. For $k = 0$ we define

$$
\alpha_0(A) \triangleq \int_A \frac{d\alpha_0}{d\varsigma}(x_0) \varsigma(dx_0), \quad A \in \mathcal{X},
$$

with, for $x_0 \in \mathsf{X}$,

$$
\frac{d\alpha_0}{d\varsigma}(x_0) \triangleq \frac{w_0(x_0) H_0^{\mathrm{u}} \ldots H_{n-1}^{\mathrm{u}}(x_0, \mathsf{X}^{n+1})}{\nu[g_0 H_0^{\mathrm{u}} \ldots H_{n-1}^{\mathrm{u}}(\cdot, \mathsf{X}^{n+1})]}.
$$

Similarly, put, for $0 \leqslant k \leqslant n-1$,

$$
(6.7) \qquad \beta_k^N(A) \triangleq \int_A \frac{d\beta_k^N}{d\tilde{\phi}_k^N}(\boldsymbol{x}_{0:k}) \tilde{\phi}_k^N(d\boldsymbol{x}_{0:k}), \quad A \in \mathcal{X}^{\otimes(k+1)},
$$

where, for $\boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}$,

$$
\frac{d\beta_k^N}{d\tilde{\phi}_k^N}(\boldsymbol{x}_{0:k}) \triangleq \frac{H_k^{\mathrm{u}} \ldots H_{n-1}^{\mathrm{u}}(\boldsymbol{x}_{0:k}, \mathsf{X}^{n+1})}{\tilde{\phi}_k^N H_k^{\mathrm{u}} \ldots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})}.
$$

The following powerful decomposition is an adaption of a similar one derived by Olsson et al. [14], Lemma 7.2 (the standard SISR case), being in turn a refinement of a decomposition originally presented by Del Moral [4].

LEMMA 6.1. *Let $n \geqslant 0$. Then, for all $f \in \mathcal{B}_{\mathrm{b}}(\mathsf{X}^{n+1})$, $N \geqslant 1$, and $r \in \{1, 2\}$,*

$$\tilde{\phi}^N_{0:n} f - \phi_n f = \sum_{k=1}^{n} A^N_k(f) + \mathbb{1}\{r = 1\} \sum_{k=0}^{n-1} B^N_k(f) + C^N(f), \tag{6.8}$$

where

$$A^N_k(f) \triangleq \frac{\sum_{i=1}^{R_N(r)} (d\alpha^N_k / d\varphi^N_k)(\tilde{\boldsymbol{\xi}}^{N,i}_{0:k}) \Psi_{k:n}[f](\tilde{\boldsymbol{\xi}}^{N,i}_{0:k})}{\sum_{j=1}^{R_N(r)} (d\alpha^N_k / d\varphi^N_k)(\tilde{\boldsymbol{\xi}}^{N,j}_{0:k})} - \alpha^N_k \Psi_{k:n}[f],$$

$$B^N_k(f) \triangleq \frac{\sum_{i=1}^{N} (d\beta^N_k / d\tilde{\phi}^N_k)(\boldsymbol{\xi}^{N,i}_{0:k}) \Psi_{k:n}[f](\boldsymbol{\xi}^{N,i}_{0:k})}{\sum_{j=1}^{N} (d\beta^N_k / d\tilde{\phi}^N_k)(\boldsymbol{\xi}^{N,j}_{0:k})} - \beta^N_k \Psi_{k:n}[f],$$

$$C^N(f) \triangleq \frac{\sum_{i=1}^{N} (d\beta_{0|n} / d\varsigma)(\boldsymbol{\xi}^{N,i}_{0}) \Psi_{0:n}[f](\boldsymbol{\xi}^{N,i}_{0})}{\sum_{j=1}^{N} (d\beta_0 / d\varsigma)(\boldsymbol{\xi}^{N,i}_{0})} - \phi_n \Psi_{0:n}[f],$$

and the operators $\Psi_{k:n} \colon \mathcal{B}_{\mathrm{b}}(\mathsf{X}^{n+1}) \to \mathcal{B}_{\mathrm{b}}(\mathsf{X}^{n+1})$, $0 \leqslant k \leqslant n$, are, for some fixed points $\hat{\boldsymbol{x}}_{0:k} \in \mathsf{X}^{k+1}$, defined by

$$\Psi_{k:n}[f] : \boldsymbol{x}_{0:k} \mapsto \frac{H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1} f(\boldsymbol{x}_{0:k})}{H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1}(\boldsymbol{x}_{0:k}, \mathsf{X}^{n+1})} - \frac{H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1} f(\hat{\boldsymbol{x}}_{0:k})}{H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1}(\hat{\boldsymbol{x}}_{0:k}, \mathsf{X}^{n+1})}.$$

Proof. Consider the decomposition

$$\begin{aligned} \tilde{\phi}^N_{0:n} f - \phi_n f = \sum_{k=1}^{n} &\left[\frac{\tilde{\phi}^N_k H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1} f}{\tilde{\phi}^N_k H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1}(\mathsf{X}^{n+1})} - \frac{\phi^N_{k-1} H^{\mathrm{u}}_{k-1} \dots H^{\mathrm{u}}_{n-1} f}{\phi^N_{k-1} H^{\mathrm{u}}_{k-1} \dots H^{\mathrm{u}}_{n-1}(\mathsf{X}^{n+1})} \right] \\ + \mathbb{1}\{r = 1\} \sum_{k=0}^{n-1} &\left[\frac{\phi^N_k H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1} f}{\phi^N_k H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1}(\mathsf{X}^{n+1})} - \frac{\tilde{\phi}^N_k H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1} f}{\tilde{\phi}^N_k H^{\mathrm{u}}_k \dots H^{\mathrm{u}}_{n-1}(\mathsf{X}^{n+1})} \right] \\ &+ \frac{\tilde{\phi}^N_0 H^{\mathrm{u}}_0 \dots H^{\mathrm{u}}_{n-1} f}{\tilde{\phi}^N_0 H^{\mathrm{u}}_0 \dots H^{\mathrm{u}}_{n-1}(\mathsf{X}^{n+1})} - \phi_n f. \end{aligned}$$

We will show that the three parts of this decomposition are identical with the three parts of (6.8). For $k \geqslant 1$, using the definitions (6.5) and (6.6) of φ^N_k and α^N_k, respectively, and following the lines of Olsson et al. [14], Lemma 7.2, we obtain

$$\begin{aligned}\frac{\phi_{k-1}^N H_{k-1}^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} H_{n-1}^{\mathrm{u}} f}{\phi_{k-1}^N H_{k-1}^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} &= \varphi_k^N \left[\frac{w_k(\cdot) H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f(\cdot)(\phi_{k-1}^N t_{k-1})}{\phi_{k-1}^N H_{k-1}^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} \right] \\ &= \alpha_k^N \left[\Psi_{k:n}[f](\cdot) + \frac{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f(\hat{\boldsymbol{x}}_{0:k})}{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\hat{\boldsymbol{x}}_{0:k}, \mathsf{X}^{n+1})} \right] \\ &= \alpha_k^N \Psi_{k:n}[f] + \frac{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f(\hat{\boldsymbol{x}}_{0:k})}{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\hat{\boldsymbol{x}}_{0:k}, \mathsf{X}^{n+1})}.\end{aligned}$$

Moreover, by definition, we get

$$\begin{aligned}&\frac{\tilde{\phi}_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f}{\tilde{\phi}_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} \\ &= \frac{\sum_{i=1}^{R_N(r)} (d\alpha_k^N / d\varphi_k^N)(\tilde{\boldsymbol{\xi}}_{0:k}^{N,i}) \Psi_{k:n}[f](\tilde{\boldsymbol{\xi}}_{0:k}^{N,i})}{\sum_{j=1}^{R_N(r)} (d\alpha_k^N / d\varphi_k^N)(\tilde{\boldsymbol{\xi}}_{0:k}^{N,j})} + \frac{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f(\hat{\boldsymbol{x}}_{0:k})}{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\hat{\boldsymbol{x}}_{0:k}, \mathsf{X}^{n+1})},\end{aligned}$$

which yields

$$\frac{\tilde{\phi}_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f}{\tilde{\phi}_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} - \frac{\phi_{k-1}^N H_{k-1}^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f}{\phi_{k-1}^N H_{k-1}^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} \equiv A_k^N(f).$$

Similarly, for $r = 1$, using the definition (6.7) of β_k^N,

$$\begin{aligned}\frac{\tilde{\phi}_{0:k}^N H_{k-1}^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f}{\tilde{\phi}_{0:k}^N H_{k-1}^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} &= \beta_k^N \left[\frac{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f(\cdot)}{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} \right] \\ &= \beta_k^N \left[\Psi_{k:n}[f](\cdot) + \frac{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f(\hat{\boldsymbol{x}}_{0:k})}{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\hat{\boldsymbol{x}}_{0:k}, \mathsf{X}^{n+1})} \right] \\ &= \beta_k^N \Psi_{k:n}[f] + \frac{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f(\hat{\boldsymbol{x}}_{0:k})}{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\hat{\boldsymbol{x}}_{0:k}, \mathsf{X}^{n+1})},\end{aligned}$$

and applying the obvious relation

$$\begin{aligned}&\frac{\phi_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f}{\phi_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} \\ &= \frac{\sum_{i=1}^{N} (d\beta_k^N / d\tilde{\phi}_k^N)(\boldsymbol{\xi}_{0:k}^{N,i}) \Psi_{k:n}[f](\boldsymbol{\xi}_{0:k}^{N,i})}{\sum_{j=1}^{N} (d\beta_k^N / d\tilde{\phi}_k^N)(\boldsymbol{\xi}_{0:k}^{N,j})} + \frac{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f(\hat{\boldsymbol{x}}_{0:k})}{H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\hat{\boldsymbol{x}}_{0:k}, \mathsf{X}^{n+1})},\end{aligned}$$

we obtain the identity

$$\frac{\phi_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f}{\phi_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} - \frac{\tilde{\phi}_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f}{\tilde{\phi}_k^N H_k^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} \equiv B_k^N(f).$$

The equality

$$\frac{\tilde{\phi}_0^N H_0^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}} f}{\tilde{\phi}_0^N H_0^{\mathrm{u}} \dots H_{n-1}^{\mathrm{u}}(\mathsf{X}^{n+1})} - \phi_n f \equiv C^N(f)$$

follows analogously. This completes the proof of the lemma. ∎

Proof of Theorem 3.3. From here on the proof is a straightforward extension of Proposition 7.1 in [14]. To establish part (i), observe the following:

- A trivial adaption of Lemmas 7.3 and 7.4 of [14] gives

$$\begin{aligned} &\|\Psi_{k:n}[f_i]\|_{\mathsf{X}^{k+1},\infty} \leqslant \operatorname{osc}(f_i)\rho^{0\vee(i-k)}, \\ &\left\|\frac{d\alpha_k^N}{d\varphi_k^N}\right\|_{\mathsf{X}^{k+1},\infty} \leqslant \frac{\|w_k\|_{\mathsf{X}^{k+1},\infty}\,\|t_{k-1}\|_{\mathsf{X}^k,\infty}}{\mu g_k(1-\rho)\epsilon_-}. \end{aligned} \tag{6.9}$$

- By mimicking the proof of Proposition 7.1 (i) in [14], that is, applying the identity $a/b - c = (a/b)(1-b) + a - c$ to each $A_k^N(f_i)$ and using twice the Marcinkiewicz–Zygmund inequality together with (6.9), we obtain the bound

$$\sqrt{R_N(r)}\|A_k^N(f_i)\|_p \leqslant B_p \frac{\operatorname{osc}(f_i)\,\|w_k\|_{\mathsf{X}^{k+1},\infty}\,\|t_{k-1}\|_{\mathsf{X}^k,\infty}}{\mu g_k(1-\rho)\epsilon_-}\rho^{0\vee(i-k)},$$

where B_p is a constant depending on p only. We refer to [14], Proposition 7.1, for details.

- For $r = 1$, inspecting the proof of Lemma 7.4 in [14] yields immediately

$$\left\|\frac{d\beta_k^N}{d\tilde{\phi}_k^N}\right\|_{\mathsf{X}^{k+1},\infty} \leqslant \frac{1}{1-\rho},$$

and repeating the arguments of the previous item for $B_k^N(f_i)$ gives

$$\sqrt{N}\|B_k^N(f_i)\|_p \leqslant B_p \frac{\operatorname{osc}(f_i)}{1-\rho}\rho^{0\vee(i-k)}.$$

- The arguments above apply directly to $C^N(f_i)$, providing

$$\sqrt{N}\|C^N(f_i)\|_p \leqslant B_p \frac{\operatorname{osc}(f_i)\,\|w_0\|_{\mathsf{X},\infty}}{\nu g_0(1-\rho)}\rho^i.$$

We conclude the proof of (i) by summing up.

The proof of (ii) (which mimics the proof of Proposition 7.1 (ii) in [14]) follows analogous lines; indeed, repeating the arguments of (i) above for the decomposition $a/b - c = (a/b)(1-b)^2 + (a-c)(1-b) + c(1-b) + a - c$ gives us the bounds

$$R_N(r)|\mathbb{E}[A_k^N(f_i)]| \leqslant B\frac{\operatorname{osc}(f_i)\,\|w_k\|^2_{\mathsf{X}^{k+1},\infty}\,\|t_{k-1}\|^2_{\mathsf{X}^k,\infty}}{(\mu g_k)^2(1-\rho)^2\epsilon_-^2}\rho^{0\vee(i-k)},$$
$$N|\mathbb{E}[B_k^N(f_i)]| \leqslant B\frac{\operatorname{osc}(f_i)}{(1-\rho)^2}\rho^{0\vee(i-k)},$$
$$N|\mathbb{E}[C^N(f_i)]| \leqslant B\frac{\operatorname{osc}(f_i)\,\|w_0\|^2_{\mathsf{X},\infty}}{(\nu g_0)^2(1-\rho)^2}\rho^{i}.$$

We refer again to [14], Proposition 7.1 (ii), for details, and summing up concludes the proof. ■

6.4. Proof of Theorem 4.1. The statement is a direct implication of Hölder's inequality. Indeed, let t_k be any first-stage importance weight function and write

$$(\phi_k t_k^*[f])^2 = \{\phi_k(t_k^{1/2}t_k^{-1/2}t_k^*[f])\}^2 \leqslant \phi_k t_k\,\phi_k\{t_k^{-1}(t_k^*[f])^2\}. \tag{6.10}$$

Now the result follows by the formula (3.3), the identity

$$\phi_k\{t_k^{-1}(t_k^*[f])^2\} = \phi_k\{t_k R_k^{\mathrm{p}}(\cdot, w_{k+1}^2\Phi_{k+1}^2[f])\},$$

and the fact that we have equality in (6.10) for $t_k = t_k^*[f]$.

Acknowledgements. The authors are grateful to Olivier Cappé who provided sensible comments on our results that improved the presentation of the paper.

REFERENCES

[1] T. Bollerslev, R. F. Engle and D. B. Nelson, *ARCH models*, in: *The Handbook of Econometrics*, Vol. 4, R. F. Engle and D. McFadden (Eds.), North-Holland, Amsterdam 1994, pp. 2959–3038.
[2] O. Cappé, É. Moulines and T. Rydén, *Inference in Hidden Markov Models*, Springer, New York 2005.
[3] N. Chopin, *Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference*, Ann. Statist. 32 (2004), pp. 2385–2411.
[4] P. Del Moral, *Feyman–Kac Formulae. Genealogical and Interacting Particle Systems with Applications*, Springer, New York 2004.
[5] R. Douc and É. Moulines, *Limit theorems for weighted samples with applications to sequential Monte Carlo methods*, Ann. Statist. 36 (2008), pp. 2344–2376.
[6] A. Doucet, N. de Freitas and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer, New York 2001.
[7] A. Doucet and A. Johansen, *A note on auxiliary particle filters*, Statist. Probab. Lett. 78 (2008), pp. 1498–1504.
[8] P. Fearnhead, *Sequential Monte Carlo Methods in Filter Theory*, Ph. D. thesis, University of Oxford, 1998.
[9] N. J. Gordon, D. J. Salmond and A. F. M. Smith, *Novel approach to non-linear/non-Gaussian Bayesian state estimation*, IEEE Proc. Comm. Radar Signal Proc. 140 (1993), pp. 107–113.

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

[11] M. Hürzeler and H. R. Künsch, *Monte Carlo approximations for general state space models*, J. Comput. Graph. Statist. 7 (1998), pp. 175–193.

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

[13] J. Liu, *Monte Carlo Strategies in Scientific Computing*,Springer, New York 2001.

[14] J. Olsson, O. Cappé, R. Douc and É. Moulines, *Sequential Monte Carlo smoothing with application to parameter estimation in non-linear state space models*, Bernoulli 14 (2008), pp. 155–179.

[15] J. Olsson, R. Douc and É. Moulines, *Improving the two-stage sampling algorithm: a statistical perspective*, in: *On Bounds and Asymptotics of Sequential Monte Carlo Methods for Filtering, Smoothing, and Maximum Likelihood Estimation in State Space Models*, Ph. D. thesis, Lund University, 2006, pp. 143–181.

[16] V. V. Petrov, *Limit Theorems of Probability Theory*, Springer, New York 1995.

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

[18] M. K. Pitt and N. Shephard, *Time varying covariances: A factor stochastic volatility approach (with discussion)*, in: *Bayesian Statistics*, Vol. 6, J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds.), Oxford University Press, Oxford 1999.

Département CITI
Télécom SudParis
9 Rue Charles Fourier, 91011 Evry Cedex, France
E-mail: randal.douc@it-sudparis.eu

Département TSI
Institut des Télécoms, Télécom ParisTech
46 Rue Barrault, 75634 Paris Cedex 13, France
E-mail: moulines@enst.fr

Center of Mathematical Sciences
Lund University
Box 118, SE-22100, Lund, Sweden
E-mail: jimmy@maths.lth.se

Received on 13.11.2007;
revised version on 3.3.2008