IMPROVING THE PERFORMANCE OF THE TWO-STAGE SAMPLING PARTICLE FILTER: A STATISTICAL PERSPECTIVE

Jimmy Olsson, Éric Moulines

Ecole Nationale Supérieure des Télécommunications
Département TSI
46 Rue Barrault, 75634 Paris Cedex 13, France

Randal Douc

Ecole Polytechnique
Centre de Mathématiques Appliquées
91128 Palaiseau cedex, France

ABSTRACT

In this paper we study asymptotic properties of weighted samples produced by the two-stage sampling (TSS) particle filter, which is a generalization of the auxiliary particle filter proposed by [1]. Besides establishing a central limit theorem (CLT) for the particle estimator of the smoothing measure, we also present bounds on the L^p error and bias of the same for a finite particle sample size. The main contribution of this article, being based on [2], is the identification of first-stage importance weights for which the increase of asymptotic variance of the CLT at a single iteration of the algorithm is minimal. Finally, we let a simple numerical example illustrate our findings.

Index Terms— Auxiliary particle filter, CLT, sequential Monte Carlo, state space models, two-stage sampling

1. INTRODUCTION

In this report a time series $Y \triangleq \{Y_k\}_{k=0}^{\infty}$, taking values in a measurable space $(\mathsf{Y}, \mathcal{Y})$, is modeled as a noisy observation of the Markov chain $X \triangleq \{X_k\}_{k=0}^{\infty}$ on some general state space $(\mathsf{X}, \mathcal{X})$. We let Q and ν be the Markov transition kernel and initial distribution of X, respectively. More specifically, the observed values of Y are conditionally independent given the hidden states of X, and the conditional distribution of Y_k depends on the state X_k only. In addition we assume that there exist, for all $x \in \mathsf{X}$, a probability density function $y \mapsto g(y, x)$ and a measure λ on $(\mathsf{Y}, \mathcal{Y})$ such that for $k \geq 0$,

$$\mathbb{P}\left(Y_n \in A | X_n = x\right) = \int_A g(y, x)\, \lambda(\mathrm{d}y) \quad \text{for } A \in \mathcal{Y}\,.$$

Furthermore we introduce, for $i \leq j$, the vector notation $\boldsymbol{X}_{i:j} \triangleq (X_i, \ldots, X_j)$; similar notation will be used for other quantities.

When operating on a *state space model* of the form described above the *smoothing distribution*

$$\phi_k(A) \triangleq \mathbb{P}\left(\boldsymbol{X}_{0:k} \in A | \boldsymbol{Y}_{0:k} = \boldsymbol{y}_{0:k}\right)\,,$$

for $A \in \mathcal{X}^{\otimes(n+1)}$, and its marginals will be highly interesting. We will throughout this paper assume that we are given a sequence $\{y_k; k \geq 0\}$ of *fixed* observations, and henceforth we let $\mathbb{P}$ and $\mathbb{E}$ denote the conditional probability measure and expectation with respect to these observations, respectively. We will also use the notation $g_k(x) \triangleq g(y_k, x)$, $x \in \mathsf{X}$. Using Bayes's formula we conclude that

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

for sets $A \in \mathcal{X}^{\otimes(k+2)}$. Nevertheless, the recursion presented above is only delusory simple since it involves the evaluation of complicated high-dimensional integrals.

Sequential Monte Carlo (SMC) methods, or *particle filters*, provide approximate solutions to the smoothing recursion (1). These methods are based on the principle of, recursively in time, approximating the smoothing distribution with the empirical distribution associated with a weighted sample.

In this article we focus on the *two-stage sampling (TSS) algorithm* suggested by [3]. The technique originates from the pioneering work by [1], who named it the *auxiliary particle filter* and proposed it as a way to prevent weight degeneracy and to robustify standard SMC methods. We provide rigorous results describing the convergence (weakly as well as in probability and L^p) of the produced approximations to the true smoothed quantities for the method in question. In addition, we discuss some possible improvements of the algorithm in the light of our findings.

2. THE TWO-STAGE SAMPLING ALGORITHM

Let us recall the TSS algorithm as presented by [3, p. 256]. In the following, denote by $\mathcal{B}_b(\mathsf{X}^m)$ the space of bounded measurable functions on X^m furnished with the supremum norm $\|f\|_{\mathsf{X}^m,\infty} \triangleq \sup_{\boldsymbol{x} \in \mathsf{X}^m} |f(\boldsymbol{x})|$. For any probability measure μ on $(\mathsf{E}, \mathcal{E})$ and μ-integrable function f, we define the expectation $\mu f \triangleq \int_{\mathsf{E}} f(x)\, \mu(\mathrm{d}x)$.

Assume that we at time k have at hand a weighted sample $\{(\boldsymbol{\xi}_k^{N,i}, \omega_k^{N,i})\}_{i=1}^N$, each $\boldsymbol{\xi}_k^{N,i}$ being referred to as a *particle*, approximating ϕ_k in the sense that $\phi_k^N f \approx \phi_k f$ for all $f \in \mathcal{B}_b(\mathsf{X}^{k+1})$, where $\phi_k^N(A) \triangleq \sum_{i=1}^N \omega_k^{N,i} \delta_{\boldsymbol{\xi}_k^{N,i}}(A)/\Omega_k^N$ and $\Omega_k^N \triangleq \sum_{i=1}^N \omega_k^{N,i}$. For $\boldsymbol{x} \in \mathsf{X}^m$, $\delta_{\boldsymbol{x}}$ here denotes the Dirac mass located at $\boldsymbol{x}$. To approximate ϕ_{k+1} we simply plug ϕ_k^N into the recursion (1) when the observation y_{k+1} becomes available, yielding for $A \in \mathcal{X}^{\otimes(k+2)}$ the mixture

$$\tilde{\phi}_{k+1}^N(A) \triangleq \sum_{i=1}^N \frac{\omega_k^{N,i} H_k^{\mathrm{u}}(\boldsymbol{\xi}_k^{N,i}, \mathsf{X}^{k+2})}{\sum_{j=1}^N \omega_k^{N,j} H_k^{\mathrm{u}}(\boldsymbol{\xi}_k^{N,j}, \mathsf{X}^{k+2})} H_k(\boldsymbol{\xi}_k^{N,i}, A) \ .$$

Here we have introduced, for $\boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}$ and $A \in \mathcal{X}^{\otimes(k+2)}$, $H_k^{\mathrm{u}}(\boldsymbol{x}_{0:k}, A) \triangleq \int_A g_{k+1}(x'_{k+1}) \, \delta_{\boldsymbol{x}_{0:k}}(\mathrm{d}\boldsymbol{x}'_{0:k}) \, Q(x_k, \mathrm{d}x'_{k+1})$, $H_k(\boldsymbol{x}_{0:k}, A) \triangleq H_k^{\mathrm{u}}(\boldsymbol{x}_{0:k}, A)/H_k^{\mathrm{u}}(\boldsymbol{x}_{0:k}, \mathsf{X}^{k+2})$. Now, since we want to form a new weighted sample estimating ϕ_{k+1}, heading for a completely recursive approximation scheme, we need to find a convenient mechanism for simulating from $\tilde{\phi}_{k+1}^N$ given the sample $\{(\boldsymbol{\xi}_k^{N,i}, \omega_k^{N,i})\}_{i=1}^N$; in fact, this is the main objective of all particle filters. In most cases cases it is possible—but generally computationally expensive—to simulate from H_k directly using auxiliary accept-reject sampling, rendering exact simulation from $\tilde{\phi}_{k+1}^N$ possible. In this case the expected number of accepted draws is however inversely proportional to $\|g_{k+1}\|_{\mathsf{X},\infty}$, which might be very large if g_{k+1} is highly peaked. A computationally cheaper solution consists in producing a weighted sample approximating $\tilde{\phi}_{k+1}^N$ by sampling from the importance sampling distribution

$$\rho_{k+1}^N(A) \triangleq \sum_{i=1}^N \frac{\omega_k^{N,i} \tau_k^{N,i}}{\sum_{j=1}^N \omega_k^{N,j} \tau_k^{N,j}} R_k^{\mathrm{p}}(\boldsymbol{\xi}_k^{N,i}, A) \ ,$$

for $A \in \mathcal{X}^{\otimes(k+2)}$. Here $\tau_k^{N,i}$, $1 \leq i \leq N$, are positive numbers referred to as *first-stage weights* and the 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 form $R_k^{\mathrm{p}}(\boldsymbol{x}_{0:k}, A) = \int_A \delta_{\boldsymbol{x}_{0:k}}(\mathrm{d}\boldsymbol{x}'_{0:k}) \, R_k(x_k, \mathrm{d}x'_{k+1})$. Thus, a draw from $R_k^{\mathrm{p}}(\boldsymbol{x}_{0:k}, \cdot)$ is produced by extending the path $\boldsymbol{x}_{0:k} \in \mathsf{X}^{k+1}$ with an additional component obtained by simulating from $R_k(x_k, \cdot)$. In this article we consider first-stage weights of type $\tau_k^{N,i} = T_k(\boldsymbol{\xi}_k^{N,i})$ for some function $T_k : \mathsf{X}^{k+1} \to \mathbb{R}^+$. Defining $W_{k+1}(\boldsymbol{x}_{0:k+1}) \triangleq [g_{k+1}(x_{k+1})/T_k(\boldsymbol{x}_{0:k})] \times \mathrm{d}Q(x_k, \cdot)/\mathrm{d}R_k(x_k, \cdot)(x_{k+1})$, $\boldsymbol{x}_{0:k+1} \in \mathsf{X}^{k+2}$, we have

$$\frac{\mathrm{d}\tilde{\phi}_{k+1}^N}{\mathrm{d}\rho_{k+1}^N}(\boldsymbol{x}_{0:k+1}) \propto \sum_{i=1}^N \mathbb{1}_{\boldsymbol{\xi}_k^{N,i}}(\boldsymbol{x}_{0:k}) W_{k+1}(\boldsymbol{x}_{0:k+1}) \ ,$$

for $\boldsymbol{x}_{0:k+1} \in \mathsf{X}^{k+2}$. A new sample $\{(\tilde{\boldsymbol{\xi}}_{k+1}^{N,i}, \tilde{\omega}_{k+1}^{N,i})\}_{i=1}^{M_N}$ targeting $\tilde{\phi}_{k+1}^N$ is hence generated by simulating M_N particles $\tilde{\boldsymbol{\xi}}_{k+1}^{N,i}$, $1 \leq i \leq M_N$, from ρ_{k+1}^N and associating with these particles the *second-stage weights* $\tilde{\omega}_{k+1}^{N,i} \triangleq W_{k+1}(\tilde{\boldsymbol{\xi}}_{k+1}^{N,i})$.

Finally, a uniformly weighted sample $\{(\boldsymbol{\xi}_{k+1}^{N,i}, 1)\}_{i=1}^N$, still targeting $\tilde{\phi}_{k+1}^N$, is obtained by resampling N of the proposed particles according to the normalized 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 recursively repeated (with $\omega_{k+1}^{N,i} = 1$, $1 \leq i \leq N$) and can be initialized by drawing $\{\xi_0^{N,i}\}_{i=1}^{M_N}$ from $\varsigma^{\otimes M_N}$, yielding $\omega_0^{N,i} = W_0(\xi_0^{N,i})$ with $W_0(x) \triangleq g_0(x) \, \mathrm{d}\nu/\mathrm{d}\varsigma(x)$, $x \in \mathsf{X}$. We summarize one step of the algorithm below.

Algorithm 1 The TSS algorithm

Ensure: $\{(\boldsymbol{\xi}_k^{N,i}, \omega_k^{N,i})\}_{i=1}^N$ approximates ϕ_k.
1: **for** $i = 1, \ldots, M_N$ **do** ▷ First stage
2: draw $I_k^{N,i}$ multinomially with respect to the normalized weights $\omega_k^{N,j}\tau_k^{N,j} / \sum_{\ell=1}^N \omega_k^{N,\ell}\tau_k^{N,\ell}$, $1 \leq j \leq N$;
3: simulate $\tilde{\boldsymbol{\xi}}_{k+1}^{N,i} \sim R_k^{\mathrm{p}}(\boldsymbol{\xi}_k^{N,I_k^{N,i}}, \cdot)$, and
4: set $\tilde{\omega}_{k+1}^{N,i} \triangleq W_{k+1}(\tilde{\boldsymbol{\xi}}_{k+1}^{N,i})$.
5: **end for**
6: **for** $i = 1, \ldots, N$ **do** ▷ Second stage
7: draw $J_{k+1}^{N,i}$ multinomially with respect to the normalized weights $\tilde{\omega}_{k+1}^{N,j} / \sum_{\ell=1}^N \tilde{\omega}_{k+1}^{N,\ell}$, $1 \leq j \leq N$, and
8: set $\boldsymbol{\xi}_{k+1}^{N,i} \triangleq \tilde{\boldsymbol{\xi}}_{k+1}^{N,J_{k+1}^{N,i}}$.
9: Finally, reset the weights: $\omega_{k+1}^{N,i} = 1$.
10: **end for**
11: Take $\{(\boldsymbol{\xi}_{k+1}^{N,i}, \omega_{k+1}^{N,i})\}_{i=1}^N$ as an approximation of ϕ_{k+1}.

The advantages of the TSS algorithm 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. It is well known that SMC methods perform well when the importance weights are well-balanced, and thus [1] propose, in the case $R_k \equiv Q$ and $\mathsf{X} = \mathbb{R}^d$, the first-stage importance weight function $T_k^{\mathrm{P\&S}}(\boldsymbol{x}_{0:k}) \triangleq g_{k+1}[\int_{\mathsf{X}} x' \, Q(x_k, \mathrm{d}x')]$. The analysis presented in the following sections will however show that this choice is not optimal in terms of minimal asymptotic (as N tends to infinity) sample variance. Moreover, theoretical results on the particle approximation stability will indicate that the second stage resampling pass should, at least for the case $M_N = N$, be cancelled, since additional resampling 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 independently in the manuscript [4]. Consequently, we advocate the scheme described in Algorithm 2. By letting $\tau_k^{N,i} \equiv 1$, $1 \leq i \leq N$, in Algorithm 2 we obtain the so-called *bootstrap particle filter*.

3. BOUNDS AND ASYMPTOTICS FOR PRODUCED APPROXIMATIONS

In [2, Theorem 3.10(i)] we establish, under the assumption that each measure $Q(x, \cdot)$, $x \in \mathsf{X}$, has a density that is uniformly bounded from below and above—pointing to appli-

Algorithm 2 Single-stage auxiliary particle filter (SSAP)

Ensure: $\{(\boldsymbol{\xi}_k^{N,i}, \omega_k^{N,i})\}_{i=1}^N$ approximates ϕ_k.
1: **for** $i = 1, \ldots, N$ **do**
2: draw $I_k^{N,i}$ multinomially with respect to the normalized weights $\omega_k^{N,j}\tau_k^{N,j} / \sum_{\ell=1}^N \omega_k^{N,\ell}\tau_k^{N,\ell}$, $1 \le j \le N$;
3: simulate $\tilde{\boldsymbol{\xi}}_{k+1}^{N,i} \sim R_k^{\mathrm{p}}(\boldsymbol{\xi}_k^{N,I_k^{N,i}}, \cdot)$, and
4: set $\omega_{k+1}^{N,i} \triangleq W_{k+1}(\tilde{\boldsymbol{\xi}}_{k+1}^{N,i})$ and $\boldsymbol{\xi}_{k+1}^{N,i} \triangleq \tilde{\boldsymbol{\xi}}_{k+1}^{N,i}$.
5: **end for**
6: Take $\{(\boldsymbol{\xi}_{k+1}^{N,i}, \omega_{k+1}^{N,i})\}_{i=1}^N$ as an approximation of ϕ_{k+1}.

cations where X is a compact set—, an L^p error bound for approximations produced by the TSS algorithm of type

$$\left\| (\tilde{\Omega}_n^N)^{-1} \sum_{j=1}^{M_N} \tilde{\omega}_n^{N,i} f_i(\tilde{\boldsymbol{\xi}}_n^{N,i}) - \phi_n f_i \right\|_p \le$$

$$\|f_i\|_{\mathsf{X}^{n+1},\infty} \left[\frac{1}{\sqrt{M_N}} \sum_{k=0}^{n} C_k \rho^{0 \vee (i-k)} + \frac{1}{\sqrt{N}} B\,(1+n-i) \right],$$

for $f_i \in \mathcal{B}_b(\mathsf{X}^{n+1})$ depending on the last $n+1-i$ states of the trajectory only. Here the constant $\rho < 1$ is the mixing rate of the hidden Markov chain when evolving conditionally on the observations. In [2, Theorem 3.10(ii)] we also state an analogous bound, this time inversely propotional to N and M_N, on the bias of the approximations. Under the mentioned assumption we may expect the model depending constants C_k, $k \ge 0$, to be roughly uniformly bounded in k. Furthermore, by inspecting the proof of this result, a similar bound is obtained for approximations produced by Algorithm 2 by simply letting $B = 0$ and $M_N = N$ in the formula above. Especially, using this latter bound for $i = n$ yields a bound on the error of the approximate *filter distribution* (that is, the marginal of ϕ_n with respect to the nth component) that is *uniformly bounded* in n. From this it is obvious that the first-stage resampling pass is enough to preserve the sample stability.

In this article, focus is however set on convergence of the approximations in the following senses.

Definition 1 (Consistency) *A sample $\{(\boldsymbol{\xi}_m^{N,i}, \omega_m^{N,i})\}_{i=1}^{M_N}$ in X^m is said to be* consistent *for the probability measure μ and the set $\mathsf{C} \subseteq \mathsf{L}^1(\mathsf{X}^m, \mu)$ if for any $f \in \mathsf{C}$, as $N \to \infty$,*

$$(\Omega_m^N)^{-1} \sum_{i=1}^{M_N} \omega_m^{N,i} f(\boldsymbol{\xi}_m^{N,i}) \xrightarrow{\mathbb{P}} \mu f\,,$$

$$(\Omega_m^N)^{-1} \max_{1 \le i \le M_N} \omega_m^{N,i} \xrightarrow{\mathbb{P}} 0\,.$$

Let μ be a probability measure on $(\mathsf{X}^m, \mathcal{X}^{\otimes m})$, γ a finite measure on $(\mathsf{X}^m, \mathcal{X}^{\otimes m})$, $\mathsf{A} \subseteq \mathsf{L}^1(\mathsf{X}^m, \mu)$ and $\mathsf{W} \subseteq \mathsf{L}^1(\mathsf{X}^m, \gamma)$ be sets of real-valued functions on X^m, and σ a real-valued non-negative functional on A.

Definition 2 (Asymptotic normality) *A weighted sample $\{(\boldsymbol{\xi}_m^{N,i}, \omega_m^{N,i})\}_{i=1}^{M_N}$ in X^m is* asymptotically normal *(abbr. a.n.) for $(\mu, \mathsf{A}, \mathsf{W}, \sigma, \gamma, \{a_N\})$ if, as $N \to \infty$,*

$$a_N(\Omega_m^N)^{-1} \sum_{i=1}^{M_N} \omega_m^{N,i}[f(\boldsymbol{\xi}_m^{N,i}) - \mu f] \xrightarrow{\mathcal{D}} \mathcal{N}[0, \sigma^2(f)]\,, f \in \mathsf{A}\,,$$

$$a_N^2(\Omega_m^N)^{-1} \sum_{i=1}^{M_N} (\omega_m^{N,i})^2 f(\boldsymbol{\xi}_m^{N,i}) \xrightarrow{\mathbb{P}} \gamma f\,, \quad f \in \mathsf{W}\,,$$

$$a_N(\Omega_m^N)^{-1} \max_{1 \le i \le M_N} \omega_m^{N,i} \xrightarrow{\mathbb{P}} 0\,.$$

The following result ([2, Theorem 3.5]) states consistency and asymptotic normality of weighted samples produced by the TSS algorithm. For brevity, we present the case $N = M_N$ only; similar results are however available also in the general case. A recent result in the same spirit has, independently of [2], been established in the manuscript [4].

Theorem 1 *Suppose that $T_k \in \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k)$ and $W_k \in \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k)$ for all $k \ge 1$, and that the equally weighted sample $\{(\xi_0^{N,i}, 1)\}_{i=1}^N$ is consistent for $[\mathsf{L}^1(\mathsf{X}, \phi_0), \phi_0]$ and a.n. for $[\phi_0, \mathsf{A}_0, \mathsf{L}^1(\mathsf{X}, \phi_0), \sigma_0, \phi_0, \{N^{1/2}\}]$. In addition, let $\mathsf{A}_0 \subseteq \mathsf{L}^1(\mathsf{X}, \phi_0)$ and define the family $\{\mathsf{A}_k\}_{k \ge 1}$ by*

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

Furthermore, let $\sigma_0 : \mathsf{A}_0 \to \mathbb{R}^+$ be a functional and define the family $\{\sigma_k\}_{k \ge 1}$ of functionals $\sigma_k : \mathsf{A}_k \to \mathbb{R}^+$ by

$$\sigma_{k+1}^2(f) \triangleq \mathrm{Var}_{\phi_{k+1}}(f) + \frac{\sigma_k^2[H_k^{\mathrm{u}}(\cdot, f - \phi_{k+1} f)]}{[\phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})]^2} + \frac{\phi_k\{T_k R_k^{\mathrm{p}}[\cdot, W_{k+1}^2 (f - \phi_{k+1} f)^2]\} \phi_k T_k}{[\phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})]^2}\,.$$

Then each sample $\{(\boldsymbol{\xi}_k^{N,i}, 1)\}_{i=1}^N$, $k \ge 1$, is consistent for $[\mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k), \phi_k]$ and a.n. for $[\phi_k, \mathsf{A}_k, \mathsf{L}^1(\mathsf{X}^{k+1}, \phi_k), \sigma_k, \phi_k, \{N^{1/2}\}]$. □

The sets $\{\mathsf{A}_k\}_{k \ge 1}$ can be controlled by imposing some additional restrictions on the model and the importance weights; indeed, if for all $k \ge 0$, $\|g_k\|_{\mathsf{X},\infty} < \infty$ and $\|W_k\|_{\mathsf{X}^{k+1},\infty} < \infty$, then it can be proved that $\mathsf{A}_0 = \mathsf{L}^2(\mathsf{X}, \phi_0)$ implies that $\mathsf{A}_k = \mathsf{L}^2(\mathsf{X}^{k+1}, \phi_k)$ for all $k \ge 1$. Moreover, by inspecting the proof of Theorem 1 one concludes that the term $\mathrm{Var}_{\phi_{k+1}}(f)$ of $\sigma_{k+1}^2(f)$ represents the cost of introducing the second-stage resampling pass, and the asymptotic variance obtained when inactivating this operation is obtained by simply *expunging the term in question from the presented formula*. Thus, bearing the stability results of SSAP algorithm in mind, there are

indeed reasons for strongly questioning whether second-stage resampling should be performed at all.

We call a first-stage importance weight function T_k *optimal* if it implies a minimal increase of asymptotic variance at a single iteration of the scheme. We should expect that such a weight involves the target function f. Indeed, define

$$T_k^*[f](\boldsymbol{x}_{0:k}) \triangleq \sqrt{\int_{\mathsf{X}} \left[\frac{\mathrm{d}H_k^{\mathrm{u}}(\boldsymbol{x}_{0:k},\cdot)}{\mathrm{d}R_k^{\mathrm{p}}(\boldsymbol{x}_{0:k},\cdot)}(\boldsymbol{x}_{0:k},x')\right]^2 \Phi_{k+1}^2[f](\boldsymbol{x}_{0:k},x')\, R_k(x_k,\mathrm{d}x')},$$

where, for $(\boldsymbol{x}_{0:k},x') \in \mathsf{X}^{k+2}$,

$$\frac{\mathrm{d}H_k^{\mathrm{u}}(\boldsymbol{x}_{0:k},\cdot)}{\mathrm{d}R_k^{\mathrm{p}}(\boldsymbol{x}_{0:k},\cdot)}(\boldsymbol{x}_{0:k},x') = g_{k+1}(x')\frac{\mathrm{d}Q(x_k,\cdot)}{\mathrm{d}R_k(x_k,\cdot)}(x') \ ,$$
$$\Phi_{k+1}[f](\boldsymbol{x}_{0:k},x') \triangleq f(\boldsymbol{x}_{0:k},x') - \phi_{k+1}f \ ,$$

and let $W_{k+1}^*[f]$ be the induced second-stage weight function; we then have the following result.

Theorem 2 *Let $k \geq 0$ and suppose that $f \in \{f' \in \mathsf{A}_{k+1} : T_k^*[f'] \in \mathsf{L}^2(\mathsf{X}^{k+1},\phi_k), W_{k+1}^*[f'] \in \mathsf{L}^1(\mathsf{X}^{k+2},\phi_{k+1})\}$. Then T_k^* is optimal and the minimal variance is given by*

$$\mathrm{Var}_{\phi_{k+1}}(f) + \frac{\sigma_k^2[H_k^{\mathrm{u}}(\cdot,f-\phi_{k+1}f)] + (\phi_k T_k^*[f])^2}{[\phi_k H_k^{\mathrm{u}}(\mathsf{X}^{k+2})]^2} \ .$$

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 $(\mathsf{E},\mathcal{E})$ and $\sum_{i=1}^d w_i = 1$, and the problem of estimating, for some given measurable and π-integrable target function f, the expectation πf. In order to relate this to the particle filtering paradigm, we will make use of the following algorithm.

Algorithm 3 Stratified importance sampling

1: **for** $k = 1,\ldots,N$ **do**
2: draw an index I_k multinomially with respect to some weights τ_i, $1 \leq i \leq d$, $\sum_{i=1}^d \tau_i = 1$;
3: simulate $\xi_k \sim \nu_{I_k}$, and
4: compute the weights $\omega_k \triangleq \frac{w_i}{\tau_i}\frac{\mathrm{d}\mu_i}{\mathrm{d}\nu_i}\Big|_{i=I_k}$
5: **end for**
6: Take $\{(\xi_k,\omega_k)\}_{k=1}^N$ as an approximation of π.

In other words, we perform Monte Carlo estimation of πf by means of sampling from some proposal mixture $\sum_{i=1}^d \tau_i\nu_i$ and forming a self-normalized estimate—cf. the technique applied in Section 2 for sampling from $\tilde{\phi}_{k+1}^N$. In this setting, the following central limit theorem can be established under weak assumptions:

$$\sqrt{N}\left[\sum_{k=1}^N \frac{\omega_k f(\xi_k)}{\sum_{\ell=1}^N \omega_\ell} - \pi f\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left[0, \sum_{i=1}^d \frac{w_i^2\alpha_i(f)}{\tau_i}\right] \ ,$$

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

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

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

$$\tau_i^* \propto w_i\alpha_i^{1/2}(f) = w_i\sqrt{\int_{\mathsf{E}} \left[\frac{\mathrm{d}\mu_i}{\mathrm{d}\nu_i}(x)\right]^2 \Pi^2[f](x)\,\nu_i(\mathrm{d}x)} \ ,$$

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.

In general the optimal weights lack closed form expressions, but in [2] it is discussed how approximations of the same can be obtained by means of a prefatory simulation pass.

Since the TSS procedure was suggested as an improvement of the bootstrap filter it is of great interest to compare the minimal asymptotic variance of Theorem 2 with the one (see e.g. [5, Theorem 6]) for the latter scheme when the same proposal kernel R_k^{p} is used in both cases. In this context we show that the minimal TSS variance is the smaller of the two if and only if

$$\mathrm{Var}_{\phi_k}\left(\{R_k^{\mathrm{p}}[\cdot,W_{k+1}^2(f-\phi_{0:k+1}f)^2]\}^{1/2}\right) \geq \mathrm{Var}_{\phi_k}\left\{R_k^{\mathrm{p}}[\cdot,W_{k+1}(f-\phi_{0:k+1}f)]\right\} \ . \quad (2)$$

Intuitively, we may expect (see [1]) that including the observations in the first-stage weights is profitable only when these are informative, whereas additional resampling for non-informative observations exclusively leads to an increase of sampling variance. The inequality (2) confirms this idea; indeed, if the likelihood g_{k+1} is highly peaked, squaring the same will amplify the peakishness, which will increase the quantity on the left hand side of (2).

4. A NUMERICAL EXAMPLE

A rigorous numerical study of the results in Section 3 is beyond the scope of this article, but in order to get an initial idea of the performace of our optimal SSAP filter we apply the method to a first order linear autoregressive process observed in noise (cf. [3, Example 7.2.3]):

$$X_{k+1} = \phi X_k + \sigma_w W_k \ ,$$
$$Y_k = X_k + \sigma_v V_k \ .$$

Here $\phi = 0.9$ and $\{W_k\}_{k=0}^\infty$ and $\{V_k\}_{k=0}^\infty$ are independent Gaussian white noise processes such that $W_k \sim \mathcal{N}(0,1)$ and

$V_k \sim \mathcal{N}(0,1)$. We let the latent chain be put at stationarity from the beginning, that is, $X_0 \sim \mathcal{N}[0, \sigma_w^2/(1-\phi^2)]$. For a linear/Gaussian model of this kind, exact expressions of the optimal weights can be obtained using the Kalman filter. In this setting we simulated, for $\sigma_v = \sigma_w = 0.1$, a record $\boldsymbol{y}_{0:10}$ of observations and estimated the filter posterior means along this trajectory by applying (1) SSAP based on true optimal allocation weights, (2) SSAP based on the generic weights $T_k^{\mathrm{P\&S}}$ of [1], and (3) the standard bootstrap filter (i.e., SSAP with $T_k \equiv 1$). Since the optimal weights are derived using asymptotic arguments we used as many as $100{,}000$ particles. The result is displayed in Figure 1(a) and it is clear that operating with true optimal allocation weights improves—in line with what we expect—the MSE performance in comparison with the other methods. The main motivation of [1] for introducing auxiliary particle filtering was to roubustify the particle approximation to outliers. Thus, we follow the lines of [3, Example 7.2.3] and repeat the experiment above for the observations $\boldsymbol{y}_{0:5} = (-0.652, -0.345, -0.676, 1.142, 0.721, 20)$, standard deviations $\sigma_v = 1$, $\sigma_w = 0.1$, and particle sample size $N = 10{,}000$. Note the large discrepancy of y_5. The outcome is plotted in Figure 1(b) from which it is evident that the optimal weights are the most efficient also in this case; moreover, the performance of the standard auxiliary particle filter is improved in comparison with the bootstrap filter. Finally, Figure 2 displays a plot of the weight functions T_4^* and $T_4^{\mathrm{P\&S}}$ for the same extreme 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 as measured by the supremum norm is still significant.

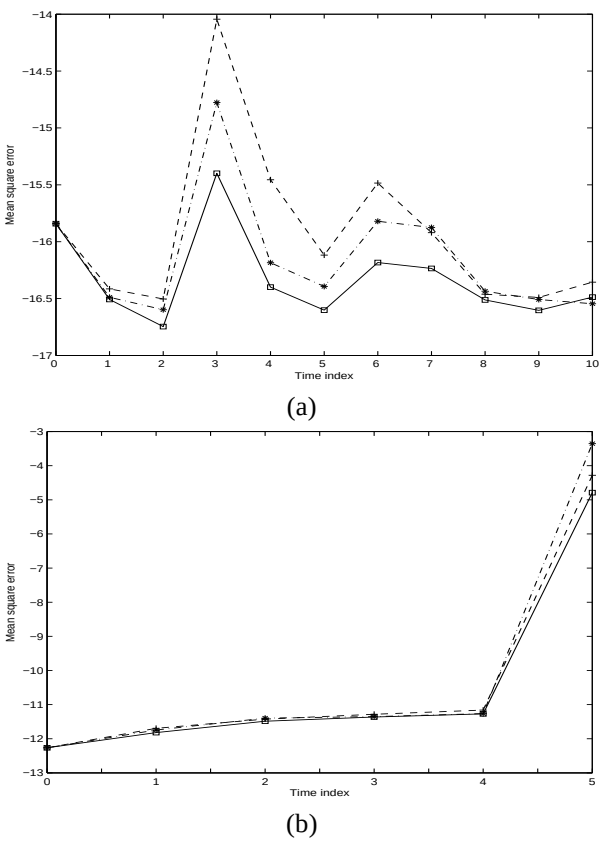

Fig. 1. Plot of MSE performances (on log-scale) of the bootstrap particle filter ($*$), SSAP based on optimal weights ($\Box$), and SSAP based on the generic weights $T_k^{\mathrm{P\&S}}$ of [1] ($+$). The MSE values are computed using 400 replications for each algorithm.

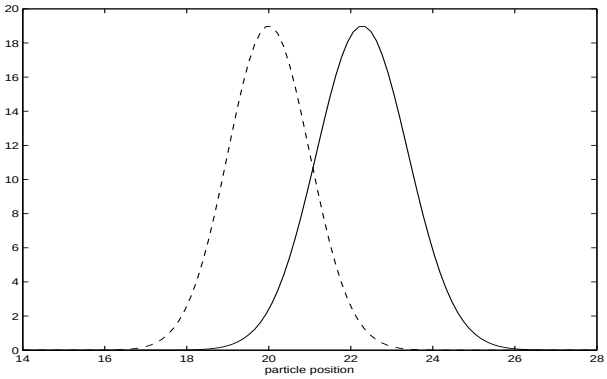

Fig. 2. Plot of the first-stage importance weight functions T_4^* (unbroken line) and $T_4^{\mathrm{P\&S}}$ (dashed line) in the presence of an outlier.

5. REFERENCES

[1] M.K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *J. Am. Statist. Assoc.*, vol. 87, pp. 493–499, 1999.

[2] R. Douc, E. Moulines, and J. Olsson, "Improving the two-stage sampling algorithm: a statistical perspective," in Doctorial Thesis 2006:13, Lund University, pp. 143–181, 2006.

[3] O. Cappé, É. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, Springer, New York, 2005.

[4] A. Doucet and A. Johansen, "A note on the auxiliary particle filter," work in progress.

[5] R. Douc and É. Moulines, "Limit theorems for weighted samples with applications to sequential monte carlo methods," *eprint* `arXiv:math.ST/0507042`, 2005.