Statistics and Computing manuscript No.
(will be inserted by the editor)

Optimal SIR algorithm vs. fully adapted auxiliary particle filter: a non asymptotical analysis

Yohan Petetin, François Desbouvries

the date of receipt and acceptance should be inserted later

Abstract Particle filters (PF) and auxiliary particle filters (APF) are widely used sequential Monte Carlo (SMC) techniques. In this paper we comparatively analyse, from a non asymptotical point of view, the Sampling Importance Resampling (SIR) PF with optimal conditional importance distribution (CID) and the fully adapted APF (FA). We compute the (finite samples) conditional second order moments of Monte Carlo (MC) estimators of a moment of interest of the filtering pdf, and analyse under which circumstances the FA-based estimator outperforms (or not) the optimal Sequential Importance Sampling (SIS)-based one. Our analysis is local, in the sense that we compare the estimators produced by one time step of the different SMC algorithms, starting from a common set of weighted points. This analysis enables us to propose a hybrid SIS/FA algorithm which automatically switches at each time step from one loop to the other. We finally validate our results via computer simulations.

Keywords Non Asymptotical Analysis, Sequential Monte Carlo, Particle Filtering, Auxiliary Particle Filtering, Sequential Importance Sampling, Resampling.

1 Introduction

Let $\{X_n\}_{n\in\mathbb{N}}$ (resp. $\{Y_n\}_{n\in\mathbb{N}}$) be hidden (resp. observed). Let $p(x_n|\mathbf{y}_{0:p})$ (or simply $p_{n|p}$), say, denote the pdf (w.r.t. Lebesgue measure) of X_n given $\mathbf{Y}_{0:p} = \mathbf{y}_{0:p}$, in which $\mathbf{Y}_{0:p} = \{Y_i\}_{i=0}^p$ and $\mathbf{y}_{0:p} = \{y_i\}_{i=0}^p$ (other pdfs of interest are defined similarly). Throughout this paper we use upper case symbols for

Telecom Institute / Telecom SudParis / CITI Department and CNRS UMR 5157, 9 rue Charles Fourier, 91011 Evry, France. Email: {yohan.petetin,francois.desbouvries}@telecom-sudparis.eu.
The final publication is available at springerlink.com.
OnLine first version: see http://www.springerlink.com/content/w1n70050770h8664/
DOI: 10.1007/s11222-012-9345-5

random variables, lower case ones for their realizations, and bold symbols for vectors; subscript indices are reserved to time indices and superscript ones to MC samples). We assume that $\{X_n, Y_n\}_{n\in\mathbb{N}}$ is a hidden Markov chain (HMC):

$$p(\mathbf{x}_{0:n}, \mathbf{y}_{0:n}) = p(x_0) \prod_{i=1}^{n} p(x_i|x_{i-1}) \prod_{i=0}^{n} p(y_i|x_i). \tag{1}$$

Bayesian filtering consists in computing the filtering pdf $p_{n|n}$ or at least some approximation of the measure $\mu_{n|n}$ with density $p_{n|n}$. As is well known $p_{n|n}$ can be computed from $p_{n-1|n-1}$ by the recursion (here $\mathcal{N}$ stands for numerator):

$$p(x_n|\mathbf{y}_{0:n}) = \frac{p(y_n|x_n) \int p(x_n|x_{n-1}) p(x_{n-1}|\mathbf{y}_{0:n-1}) \mathrm{d}x_{n-1}}{p(y_n|\mathbf{y}_{0:n-1}) = \int \mathcal{N} \mathrm{d}x_n}. \tag{2}$$

In general (2) cannot be computed exactly, and one needs to resort to approximations. Among them PF [1] [2] [3] [4] and APF algorithms [5] [6] [7] are SMC methods which propagate a discrete approximation $\widehat{\mu}_{n|n}$ of $\mu_{n|n}$.

From a theoretical point of view, convergence results of some SMC algorithms have been proven in a series of papers, see e.g. [8] [9] [10] [11] and references therein. These results have been recently extended to the APF [12], and a tuning of the first-stage weights of the APF from an asymptotical variance analysis viewpoint has also been proposed in [13].

All these results are asymptotical, i.e. they hold if the number of particles tends to infinity. In this paper, we propose a *non asymptotical* comparative analysis of two SMC algorithms: the PF with optimal CID [14], without (SIS) or with (SIR) resampling, versus the FA algorithm [5]. On the other hand, our analysis is local in the sense that we compare the samples produced by one time step of the SIR, SIS and FA algorithms, given a common set of samples at time $n-1$. So let

$$\Theta_n = \mathrm{E}_{p_{n|n}}(f(X)) \tag{3}$$

be a moment of $p_{n|n}$ (here $f(.)$ is some function of interest, for which we assume that Θ_n exists), and let $\widehat{\Theta}_n^{\mathrm{FA}}$, $\widehat{\Theta}_n^{\mathrm{SIS}}$ and $\widehat{\Theta}_n^{\mathrm{SIR}}$ be MC estimates of Θ_n computed from the FA, SIS or SIR algorithms. Starting from a discrete approximation $\widehat{\mu}_{n-1|n-1} = \sum_{i=1}^{N} w_{n-1}^i \delta_{x_{n-1}^i}$ of $\mu_{n-1|n-1}$ (δ_x is the Dirac mass at point x), $\widehat{\Theta}_n^{\mathrm{FA}}$, $\widehat{\Theta}_n^{\mathrm{SIS}}$ and $\widehat{\Theta}_n^{\mathrm{SIR}}$ share the same mean, but the variance of $\widehat{\Theta}_n^{\mathrm{SIR}}$ always exceeds that of $\widehat{\Theta}_n^{\mathrm{FA}}$ and that of $\widehat{\Theta}_n^{\mathrm{SIS}}$. On the other hand, none of the FA- and SIS-based estimators systematically outperforms the other; as we shall see, it is nevertheless possible to analyse the comparative behavior of both algorithms in terms of relevant parameters of the underlying model.

This paper is organized as follows. In §2 we briefly recall the optimal SIR and FA algorithms. In §3 we compute, for an arbitrary finite number of particles, the conditional second order moments of $\widehat{\Theta}_n^{\mathrm{FA}}$, $\widehat{\Theta}_n^{\mathrm{SIS}}$ and $\widehat{\Theta}_n^{\mathrm{SIR}}$. In §4 we analyse these results and we enlighten the role played in the variance of the estimators by the number of efficient particles $N_{eff}(n)$; from these results we propose a hybrid SIS/FA algorithm which at each time step switches between the SIS loop or the FA one, depending on local conditions. Finally in

§5 we compare our algorithms via simulations, and we end the paper with a conclusion.

2 Optimal SIR and FA algorithms

Let us first recall the SIR algorithm with optimal CID and the FA algorithm.

- PF algorithms (see e.g. [3] [4]) are based on importance sampling (IS). Let $\mu_{0:n|n}$ be the measure with density $p(\mathbf{x}_{0:n}|\mathbf{y}_{0:n})$. Assume that $\mu_{0:n-1|n-1}$ is approximated by $\widehat{\mu}_{0:n-1|n-1} = \sum_{i=1}^N w_{n-1}^i \delta_{\mathbf{x}_{0:n-1}^i}$. At time n we sample x_n^i from some CID $q(x_n|\mathbf{x}_{0:n-1}^i, \mathbf{y}_{0:n})$ and update w_{n-1}^i as $w_n^i \propto \frac{p(x_n^i, y_n|x_{n-1}^i)}{q(x_n^i|\mathbf{x}_{0:n-1}^i, \mathbf{y}_{0:n})}$ $\times w_{n-1}^i$, $\sum_{i=1}^N w_n^i = 1$. Then $\sum_{i=1}^N w_n^i \delta_{\mathbf{x}_{0:n}^i}$ is an MC approximation of $\mu_{0:n|n}$, and thus $\sum_{i=1}^N w_n^i \delta_{x_n^i}$ is an MC approximation of $\mu_{n|n}$.
  Now such SIS algorithms are well known to suffer from weights degeneracy. Two main rescues are available. First, it has proved important to resample from $\sum_{i=1}^N w_n^i \delta_{x_n^i}$, either systematically or according to some strategy, and variants are available, see e.g. [15] [16] [17]. The second important point is to choose the CID carefully. It has proved of interest [18] [19] [20] [14], if possible, to sample the particles from the optimal CID $q^{opt}(x_n|\mathbf{x}_{0:n-1}^i, \mathbf{y}_{0:n}) = p(x_n|x_{n-1}^i, y_n)$, i.e. the distribution which minimizes the variance of the importance weights w_n^i, given observations $\mathbf{y}_{0:n}$ and past samples $\mathbf{x}_{0:n-1}^i$. This minimum of the conditional variance is zero, and indeed for this choice of the CID, $w_n^i \propto p(y_n|x_{n-1}^i) w_{n-1}^i$, which means that the new weight w_n^i does not depend on the particle x_n^i which has just been sampled.
- On the other hand, the rationale of APF algorithms [5] [6] (see also [11, §8.1] and [12] [13] [7] for recent developments) is as follows. If exact computing of (2) is not possible, one can plug into (2) some discrete approximation $\sum_{i=1}^N w_{n-1}^i \ \delta_{x_{n-1}^i}$ of $\mu_{n-1|n-1}$, which yields a mixture pdf approximation $\pi_{n|n}$ of $p_{n|n}$:

$$\pi_{n|n}(x_n) = \sum_{i=1}^N \ w_n^i p(x_n|x_{n-1}^i, y_n), \tag{4}$$

$$w_n^i \propto p(y_n|x_{n-1}^i) w_{n-1}^i, \sum_{i=1}^N w_n^i = 1. \tag{5}$$

  Next one should get (approximate) samples from (4). If sampling from $\pi_{n|n}$ is impossible, APF algorithms are IS based SMC techniques which target $\pi_{n|n}$ via another mixture pdf. Now, if direct sampling from $\pi_{n|n}$ is indeed feasible we are in the so-called FA case, in the terminology of [5].

We now summarize both the SIS/SIR algorithm with optimal CID, and the FA algorithm. In the SIS algorithm we sample (S) points $x_n^{\mathrm{SIS},i}$ and update

(W) weigths w_n^i, or vice versa, since in the optimal case both steps can be interverted; the SIR algorithm adds an optional resampling (R) step. In the FA algorithm we sample from mixture $\pi_{n|n}$; using the composition method (see e.g. [21]), this amounts to computing weights w_n^i (W step), and next sampling $\widetilde{x}_{n-1}^i$ from the discrete measure $\sum_{i=1}^N w_n^i \delta_{x_{n-1}^i}$ (R step), then $x_n^{\mathrm{FA},i}$ from the continuous pdf $p(x_n|\widetilde{x}_{n-1}^i, y_n)$ which has been selected (S step). In summary:[1]

Optimal SIS/SIR algorithm vs. FA algorithm.

Let $\widehat{\mu}_{n-1|n-1} = \sum_{i=1}^N w_{n-1}^i \delta_{x_{n-1}^i}$ be an approximation of $\mu_{n-1|n-1}$.

W. For $1 \leq i \leq N$, compute $w_n^i \propto p(y_n|x_{n-1}^i) w_{n-1}^i$, $\sum_{i=1}^N w_n^i = 1$; then

$\rightarrow$ *if SIS/SIR algorithm with optimal CID:*

S. For $1 \leq i \leq N$, sample $x_n^{\mathrm{SIS},i}$ from $p(x_n|x_{n-1}^i, y_n)$;
R. For $1 \leq i \leq N'$, sample $x_n^{\mathrm{SIR},i} \sim \sum_{i=1}^N w_n^i \delta_{x_n^{\mathrm{SIS},i}}$.

$\rightarrow$ *if FA algorithm:*

R. For $1 \leq i \leq N'$, sample $\widetilde{x}_{n-1}^i \sim \sum_{i=1}^N w_n^i \delta_{x_{n-1}^i}$;
S. For $1 \leq i \leq N'$, sample $x_n^{\mathrm{FA},i}$ from $p(x_n|\widetilde{x}_{n-1}^i, y_n)$.

Then $\sum_{i=1}^N w_n^i \delta_{x_n^{\mathrm{SIS},i}}$, $\sum_{i=1}^{N'} \frac{1}{N'} \delta_{x_n^{\mathrm{SIR},i}}$ or $\sum_{i=1}^{N'} \frac{1}{N'} \delta_{x_n^{\mathrm{FA},i}}$ approximate $\mu_{n|n}$, and $\widehat{\Theta}_n^{\mathrm{SIS}} = \sum_{i=1}^N w_n^i f(X_n^{\mathrm{SIS},i})$, $\widehat{\Theta}_n^{\mathrm{SIR},N'} = \frac{1}{N'} \sum_{i=1}^{N'} f(X_n^{\mathrm{SIR},i})$ and $\widehat{\Theta}_n^{\mathrm{FA},N'} = \frac{1}{N'} \times \sum_{i=1}^{N'} f(X_n^{\mathrm{FA},i})$ are three candidate estimators of moment Θ_n defined in (3).

3 A comparative analysis of the SIS, SIR and FA algorithms

As we see the SIR and FA algorithms share the same operations but in a different order: in the SIR algorithm the successive steps are $W \rightarrow S \rightarrow R$, and in the FA one $W \rightarrow R \rightarrow S$. Let us now comparatively analyse these algorithms, i.e. measure the effect of resampling before or after the sampling step.

We begin with a simple observation. As was recalled in §2, in an SIS algorithm weigths degenerate with time, so a lot of computational effort is devoted to sampling particles which will be associated to vanishing weights. Resampling was introduced as a rescue to this problem, and indeed this technique eliminates particles associated with weights that are too low. But since in the SIR algorithm a point $x_n^{\mathrm{SIR},i}$ can be sampled several times, $\{x_n^{\mathrm{SIR},i}\}_{i=1}^{N'}$ is a subset of $\{x_n^{\mathrm{SIS},i}\}_{i=1}^{N}$. So the support of the distribution shrinks, which is another form of degeneracy. Quantitatively this support degeneracy can be measured by the (conditional) average size of the support after resampling $\mathrm{E}(\nu')$, defined

[1] In view of the analysis of sections 3 and 4 we consider the case where the number of samples N' at time n is possibly different to that N at time $n-1$, even though in SMC techniques we often take $N' = N$.

as the expectation of the number of points which have been (re-)selected at least once. Since the total number of outcomes of each particle after the R step follows a multinomial distribution with parameters w_n^i,

$$\mathrm{E}(\nu') = N - \sum_{i=1}^{N} (1 - w_n^i)^{N'}. \tag{6}$$

By contrast in the FA algorithm the support cannot degenerate. Of course, the R step still induces support shrinkage, so $\{\widetilde{x}_{n-1}^i\}_{i=1}^{N'}$ is a subset of $\{x_{n-1}^i\}_{i=1}^{N}$; but next each $x_n^{\mathrm{FA},i}$ is drawn from pdf $p(x_n^i|\widetilde{x}_{n-1}^i, y_n)$, so particles $x_n^{\mathrm{FA},i}$ and $x_n^{\mathrm{FA},j}$ will be different even if $\widetilde{x}_{n-1}^i = \widetilde{x}_{n-1}^j$.

We next turn to statistical considerations. Let us first compare the samples produced by one time step of both algorithms. By construction, $\{x_n^{\mathrm{FA},i}\}_{i=1}^{N'}$ are i.i.d. from $\pi_{n|n}$; since $\pi_{n|n}$ is a mixture $\sum_{k=1}^{N} \alpha_k p_k(x)$, each $x_n^{\mathrm{FA},i}$ is actually produced by drawing an index j from $\sum_{k=1}^{N} \alpha_k \delta_k$, and next a sample from the $p_j(x)$ which has been selected. By contrast, in the SIR algorithm these discrete and continous parts are interverted: one first draws one sample from each continuous pdf $p_i(x)$, and next reselects them according to $\sum_{i=1}^{N} \alpha_i \delta_i$. As a consequence [22], given $\{(x_{n-1}^i, w_{n-1}^i)\}_{i=1}^{N}$ and y_n, $\{x_n^{\mathrm{SIR},i}\}_{i=1}^{N'}$ remain identically distributed from $\pi_{n|n}$ but are no longer independent. (Of course, by definition of the resampling step in PF algorithms, $\{x_n^{\mathrm{SIR},i}\}_{i=1}^{N'}$ are independent given $\{(x_{n-1}^i, w_{n-1}^i)\}_{i=1}^{N}$, y_n *and* $\{x_n^{\mathrm{SIS},i}\}_{i=1}^{N}$; they become dependent (given $\{(x_{n-1}^i, w_{n-1}^i)\}_{i=1}^{N}$ and y_n) if the conditioning on $\{x_n^{\mathrm{SIS},i}\}_{i=1}^{N}$ is removed.)

We next compare estimators $\widehat{\Theta}_n^{\mathrm{SIS}}$, $\widehat{\Theta}_n^{\mathrm{SIR},N'}$ and $\widehat{\Theta}_n^{\mathrm{FA},N'}$ of Θ_n:

Theorem 1 *Let $\widehat{\Theta}_n^{\mathrm{SIS}}$, $\widehat{\Theta}_n^{\mathrm{SIR},N'}$ and $\widehat{\Theta}_n^{\mathrm{FA},N'}$ be defined as above. Let $\widehat{\Theta}_n^{\mathrm{SIR},N} = \widehat{\Theta}_n^{\mathrm{SIR}}$, $\widehat{\Theta}_n^{\mathrm{FA},N} = \widehat{\Theta}_n^{\mathrm{FA}}$ and $\pi_{n|n}(.)$ be given by* (4)*. Then* $\mathrm{E}(\widehat{\Theta}_n^{\mathrm{FA},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) = \mathrm{E}(\widehat{\Theta}_n^{\mathrm{SIS}}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) = \mathrm{E}(\widehat{\Theta}_n^{\mathrm{SIR},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) = \mathrm{E}_{\pi_{n|n}}(f(X))$, *and*

$$\mathrm{var}(\widehat{\Theta}_n^{\mathrm{FA},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) = \frac{1}{N'} \mathrm{var}_{\pi_{n|n}}(f(X)), \tag{7}$$

$$\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS}}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) = \sum_{i=1}^{N} (w_n^i)^2 \mathrm{var}_{p(x_n|x_{n-1}^i, y_n)}(f(X)), \tag{8}$$

$$\begin{aligned} \mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIR},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) = {} & \frac{1}{N'} \mathrm{var}_{\pi_{n|n}}(f(X)) \\ & + \frac{N'-1}{N'} \sum_{i=1}^{N} (w_n^i)^2 \mathrm{var}_{p(x_n|x_{n-1}^i, y_n)}(f(X)). \end{aligned} \tag{9}$$

Proof For notational simplicity we drop the conditioning on $\mathbf{x}_{n-1}^{1:N}$, $\mathbf{w}_{n-1}^{1:N}$ and y_n. Only (9) is not straightforward, so let us compute

$$\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIR},N'}) = \frac{1}{N'^2} \sum_{i=1}^{N'} \mathrm{var}(f(X_n^{\mathrm{SIR},i})) + \frac{1}{N'^2} \sum_{\substack{k,l=1 \\ k \neq l}}^{N'} \mathrm{Cov}(f(X_n^{\mathrm{SIR},k}), f(X_n^{\mathrm{SIR},l})).$$

Since $X_n^{\mathrm{SIR},i} \sim \pi_{n|n}(.)$, the first term is equal to $\frac{1}{N'}\mathrm{var}_{\pi_{n|n}}(f(X))$. Next

$$\mathrm{E}(f(X_n^{\mathrm{SIR},k})|\mathbf{x}_n^{\mathrm{SIS},1:N}) = \widehat{\Theta}_n^{\mathrm{SIS}}(\mathbf{x}_n^{\mathrm{SIS},1:N}), \tag{10}$$

and from (10), for all k, l, $1 \leq k,l, \leq N'$ with $k \neq l$, $\mathrm{E}(f(X_n^{\mathrm{SIR},k})f(X_n^{\mathrm{SIR},l})|$ $\mathbf{x}_n^{\mathrm{SIS},1:N}) = (\widehat{\Theta}_n^{\mathrm{SIS}}(\mathbf{x}_n^{\mathrm{SIS},1:N}))^2$, so $\mathrm{E}(f(X_n^{\mathrm{SIR},k})f(X_n^{\mathrm{SIR},l})) = \mathrm{E}(\widehat{\Theta}_n^{\mathrm{SIS}})^2$, and finally $\mathrm{Cov}(f(X_n^{\mathrm{SIR},k}), f(X_n^{\mathrm{SIR},l})) = \mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS}})$, whence (9). □

4 Discussion

We now comment Theorem 1. Given $\{x_{n-1}^i, w_{n-1}^i\}_{i=1}^N$ and y_n, $\widehat{\Theta}_n^{\mathrm{FA},N'}$, $\widehat{\Theta}_n^{\mathrm{SIS}}$ and $\widehat{\Theta}_n^{\mathrm{SIR},N'}$ have the same mean, so we compare their conditional variances.

4.1 SIR- vs. FA-based estimators

From (7) and (9) we see that for all N and N',

$$\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIR},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) > \mathrm{var}(\widehat{\Theta}_n^{\mathrm{FA},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n). \tag{11}$$

Starting from a common empirical measure $\widehat{\mu}_{n-1|n-1}$, and after the weights have been updated, it is thus preferable to resample and next sample, rather than sampling first and next resampling.

Remark 1 Inequality (11) is easily understood if we rewrite the optimal SIR algorithm. In the optimal case, the succession of the S, W and R steps is equivalent to that of the W, S and R ones, which led to the formulation given at the end of section 2. But since the R step does not depend on the particles $\{x_n^{\mathrm{SIS},i}\}_{i=1}^N$ which have just been sampled, it can be performed before the S step, provided we sample new particles only when necessary; so the SIR algorithm can again be rewritten as follows:

W. For $1 \leq j \leq N$, compute $w_n^j \propto p(y_n|x_{n-1}^j)w_{n-1}^j$, $\sum_{j=1}^N w_n^j = 1$;
R. For $1 \leq j \leq N'$, sample index $i_j \sim \sum_{i=1}^N w_n^i \delta_i$;
S. For $1 \leq j \leq N'$,
- If $i_j \notin \{i_1, \cdots, i_{j-1}\}$ then sample $x_n^{\mathrm{SIS},j} \sim p(x_n|x_{n-1}^{i_j}, y_n)$, and set $(x_n^{\mathrm{SIR},j}, w_n^j) = (x_n^{\mathrm{SIS},j}, 1/N')$;
- If $i_j = i_k$ for $i_k \in \{i_1, \cdots, i_{j-1}\}$ then set $(x_n^{\mathrm{SIR},j}, w_n^j) = (x_n^{\mathrm{SIR},k}, 1/N')$.

This implementation of the SIR algorithm is computationally more efficient than the original one since we no longer draw particles $x_n^{\mathrm{SIS},i}$ which finally will be discarded at the resampling step. More importantly, it enlightens the lack of diversity of the SIR algorithm which results in (11): in the SIR algorithm a given index i_j is represented by one single particle, while in the FA algorithm it is represented by N_{i_j} ones, where N_{i_j} is the number of outcomes of i_j.

More precisely, this lack of diversity (measured by the second term $\delta(N')$ of the r.h.s. of (9)) depends on parameters $\{\mathrm{var}_{p(x_n|x_{n-1}^i, y_n)}(f(X))\}_{i=1}^N$ and on

the so-called number of efficient particles $N_{eff}(n)$ (an estimator of the effective sample size) after updating of the weights, introduced in [19] (see also [20], [23] and [24]) and defined as

$$N_{eff}(n) \stackrel{\text{def}}{=} \frac{1}{\sum_{i=1}^{N}(w_n^i)^2}. \tag{12}$$

If $\text{var}_{p(x_n|x_{n-1}^i,y_n)}(f(X))$ is small for all i then the N_{i_j} points $f(x_n^{\text{FA},k})$ associated to index i_j almost reduce to a single one, and $\delta(N')$ is small. If $\text{var}_{p(x_n|x_{n-1}^i,y_n)}(f(X))$ is almost constant (equal to γ) then $\delta(N') \approx \frac{\gamma}{N_{eff}(n)}$. So $\delta(N')$ increases as $N_{eff}(n)$ decreases, and indeed when $N_{eff}(n)$ is small, the SIR algorithm produces very few different particles $x_n^{\text{SIR},k}$ while the FA one still produces N' different ones. But again the effect of these N' samples is unsignificant if $f(x_n^{\text{FA},k})$ are not well separated, i.e. if numerator γ is small.

Remark 2 (11) holds for all N and N'. Note however that in order to get $N' = N + K$ samples from $\pi_{n|n}$, the SIR procedure computes $2N + K$ samples and the FA one $2N + 2K$ ones. It is thus of interest to compare both estimators *at fixed cost*, i.e. for an overall fixed number of samples $2N+2K$; we thus compare $\widehat{\Theta}_n^{\text{FA},N+K}$ to $\widehat{\Theta}_n^{\text{SIR},N+2K}$, for some fixed value of N and for $K > -N/2$. So let

$$\begin{aligned}\delta'(K) &\stackrel{\text{def}}{=} \text{var}(\widehat{\Theta}_n^{\text{SIR},N+2K}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) - \text{var}(\widehat{\Theta}_n^{\text{FA},N+K}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) \\ &= \frac{\text{var}(\widehat{\Theta}_n^{\text{FA}}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n)}{(N+2K)(N+K)} \times a(K)\end{aligned}$$

in which $a(K) = -NK+(N+K)(N+2K-1)A_n$ and $A_n = \frac{\text{var}(\widehat{\Theta}_n^{\text{SIS}}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n)}{\text{var}(\widehat{\Theta}_n^{\text{FA}}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n)}$. The sign of $\delta'(K)$ is equal to that of polynomial $a(K)$, which is positive if $-\frac{N}{2} < K \leq 0$. Let now $K > 0$. If $A_n \geq 1$ then $a(K) \geq N^2+(N+K)(2K-1) \geq 0$. Finally, for any $N > 1$ and $K > -\frac{N}{2}$, $a(K)$ and thus $\delta'(K)$ is negative iff. $A_n \leq \alpha_1(N) = \frac{N(3N-1)-2N(2N(N-1))^{\frac{1}{2}}}{(N+1)^2}$ and $K_1(A_n, N) \leq K \leq K_2(A_n, N)$, in which $K_i(A_n, N) = \frac{-A_n(3N-1)+N+\epsilon_i(A_n^2(N+1)^2-2A_nN(3N-1)+N^2)^{\frac{1}{2}}}{4A_n}$, $\epsilon_1 = 1$ and $\epsilon_2 = -1$.

Let us comment this result. First, note that when N is high, $\alpha_1(N) \approx 3 - \sqrt{8} \approx 0.1716$. It means that $\delta'(K)$ can be negative only if the N samples SIS-based estimator widely outperforms the FA-based one. Experimentally, this is a strong condition but can happen nevertheless, see Figure 6 in section 5.2. Let us take for instance $N = 1000$ and $A_n = 0.16$. Then $K_1(A, N) \approx 412$ and $K_2(A, N) \approx 1214$, i.e. $\widehat{\Theta}_n^{\text{SIR},2K}$ outperforms $\widehat{\Theta}_n^{\text{FA},K}$ only for very particular choices of K.

Remark 3 Let $J_n^{N'} = \text{E}((\widehat{\Theta}_n^{N'} - \Theta_n)^2|\mathbf{y}_{0:n})$ be the MSE (conditionally on the observations) of $\widehat{\Theta}_n^{N'}$, where $\widehat{\Theta}_n^{N'}$ here represents either $\widehat{\Theta}_n^{\text{FA},N'}$ or $\widehat{\Theta}_n^{\text{SIR},N'}$. From Theorem 1, $\text{E}((\widehat{\Theta}_n^{N'}-\Theta_n)^2|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, \mathbf{y}_{0:n}) = \text{var}(\widehat{\Theta}_n^{N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, \mathbf{y}_n)$ $+(\text{E}_{\pi_{n|n}}(f(X)) - \Theta_n)^2$. Taking expectations on both sides we see that

$$J_n^{N'} = \text{E}(\text{var}(\widehat{\Theta}_n^{N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, \mathbf{y}_{0:n})|\mathbf{y}_{0:n}) + \text{E}([\text{E}_{\pi_{n|n}}(f(X)) - \Theta_n]^2|\mathbf{y}_{0:n})$$

Since the second term is constant, from (11) we get $J_n^{\mathrm{SIR},N'} > J_n^{\mathrm{FA},N'}$ as well.

4.2 SIR- vs. SIS-based estimators

Even though (9) decomposes $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIR},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n)$ as the sum of two positive terms, it should not be confused with the Rao-Blackwell (RB) equality which in the context of this paper reads

$$\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIR},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) = \mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS}}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) + \Delta_n, \quad (13)$$

in which

$$\Delta_n = \mathrm{E}((\widehat{\Theta}_n^{\mathrm{SIR},N'} - \widehat{\Theta}_n^{\mathrm{SIS}})^2|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) \geq 0 \quad (14)$$

is an extra variance term which is due to the resampling step only. Equation (13) enables to conclude, as is well known [20, Section 4.2.1] [11, p. 213], that

$$\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIR},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) \geq \mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS}}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) \quad (15)$$

(and so, as above, $J_n^{\mathrm{SIR},N'} > J_n^{\mathrm{SIS}}$ as well). So whatever N' (see however Remark 5 below), it is preferable to compute $\widehat{\Theta}_n$ from the set of weighted points $\{(x_n^{\mathrm{SIS},i}, w_n^i)\}_{i=1}^N$ rather than from the set of unweighted ones $\{(x_n^{\mathrm{SIR},i}, 1/N')\}_{i=1}^{N'}$: in addition to introducing dependency, support shrinkage and extra computational cost, resampling also induces extra estimator variance.

Now, resampling was designed as a rescue against weights degeneracy, and indeed there is no drawback in discarding particles which are associated to very low weights. However if the entropy of the weights gets higher (i.e. all weights become almost equal) then the risk of resampling is to eliminate good particles. Heuristically we thus expect that the larger $N_{eff}(n)$ (defined in (12)) is, the larger the extra variance term Δ_n is. This is not true stricto senso, but as we now see, Δ_n somehow reflects and measures whether resampling should indeed be done or not. Two arguments are available.

4.2.1 Resampling vs. the risk of discarding good particles.

Let $N' = N$. Consider the particular case where only $\widetilde{N}_{eff}$ particles have non-null weights, and the total mass is distributed uniformly on these $\widetilde{N}_{eff}$ particles. Then (6) becomes $\mathrm{E}(\nu') = \widetilde{N}_{eff} - \widetilde{N}_{eff}(1 - \frac{1}{\widetilde{N}_{eff}})^N$. Let $h(k) = k(1 - \frac{1}{k})^N$. Function h increases with k, so the (average) number of particles discarded by the resampling step increases as $\widetilde{N}_{eff}$ increases. Function $\mathrm{E}(\nu')$ is plotted on Figure 1.

- If $\widetilde{N}_{eff} = 1$, then we are sure to resample the particle, and the moment estimators $\widehat{\Theta}_n^{\mathrm{SIR}}$ and $\widehat{\Theta}_n^{\mathrm{SIS}}$ coincide;
- $h'(.)$ has an inflexion point at $\widetilde{N}_{eff} = \frac{N+1}{3}$: as long as $\widetilde{N}_{eff} \leq \frac{N+1}{3}$ the average number of lost particles $h(\widetilde{N}_{eff})$ is weak, but becomes larger after this point;

– If $\widetilde{N}_{eff} = N$ then $\mathrm{E}(\nu') \approx \frac{2N}{3}$, which means that we lost on average one third of the particles.

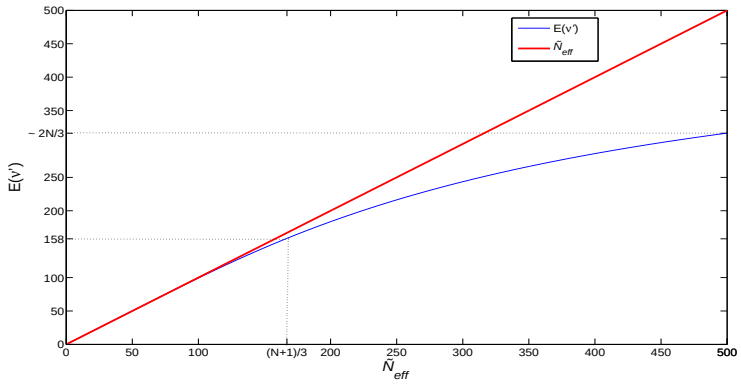

Fig. 1 Average size of the support after resampling, in function of $\widetilde{N}_{eff}$. Support shrinkage becomes severe when $\widetilde{N}_{eff}$ gets close to N (here $N = 500$).

4.2.2 Role of $N_{eff}(n)$ in the extra variance term Δ_n.

In situations where particles are (almost) equally weighted the support shrinkage is severe; we expect in such situations that Δ_n becomes significant. By contrast if we run an SIS algorithm without resampling, weights will degenerate with time, $N_{eff}(n)$ will decrease and we expect Δ_n to decrease too, because in such a case resampling would indeed be beneficial. Of course, Δ_n does not depend just on $N_{eff}(n)$ but also on model (1) and on function f. Let us first rewrite Δ_n as

$$\Delta_n = A_n + B_n, \tag{16}$$

$$A_n = \sum_{i=1}^{N} \frac{1}{N'} w_n^i (1 - w_n^i) \mathrm{var}_{p(x_n|x_{n-1}^i, y_n)}(f(X)), \tag{17}$$

$$B_n = \frac{1}{N'} \sum_{1 \le i < j \le N} w_n^i w_n^j \left[\mathrm{E}_{p(x_n|x_{n-1}^i, y_n)}(f(X)) - \mathrm{E}_{p(x_n|x_{n-1}^j, y_n)}(f(X)) \right]^2. \tag{18}$$

Proof From equations (8) and (9), $\Delta_n = \frac{1}{N'} \mathrm{var}_{\pi_{n|n}}(f(X)) - \frac{1}{N'} \sum_{i=1}^{N} (w_n^i)^2 \times \mathrm{var}_{p(x_n|x_{n-1}^i, y_n)}(f(X))$. From (4), $\mathrm{var}_{\pi_{n|n}}(f(X)) = \sum_{i=1}^{N} w_n^i [\mathrm{var}_{p(x_n|x_{n-1}^i, y_n)}(f(X)) + (\mathrm{E}_{p(x_n|x_{n-1}^i, y_n)}(f(X)))^2] - [\sum_{i=1}^{N} w_n^i \mathrm{E}_{p(x_n|x_{n-1}^i, y_n)}(f(X))]^2$. So $\Delta_n = A_n + B_n$, in which A_n is given by (17), and

$$B_n = \frac{1}{N'} \left(\sum_{i=1}^{N} w_n^i \left[\mathrm{E}_{p(x_n|x_{n-1}^i, y_n)}(f(X)) \right]^2 - \left[\sum_{i=1}^{N} w_n^i \mathrm{E}_{p(x_n|x_{n-1}^i, y_n)}(f(X)) \right]^2 \right) \tag{19}$$

Multiplying the first term of (19) by $\sum_{j=1}^{N} w_n^j$ (which is equal to 1), and rearranging the terms of the sum, we get (18). □

As we now see, the variances in (17) and square differences of the means in (18) are the parameters which play an important role in the variation of Δ_n:

Proposition 1 *Let*

$$p = \min_{i=1\cdots N} \mathrm{var}_{p(x_n|x_{n-1}^i,y_n)}(f(X)), \tag{20}$$

$$P = \max_{i=1\cdots N} \mathrm{var}_{p(x_n|x_{n-1}^i,y_n)}(f(X)), \tag{21}$$

$$m = \min_{1\leq i<j\leq N} [\mathrm{E}_{p(x_n|x_{n-1}^i,y_n)}(f(X)) - \mathrm{E}_{p(x_n|x_{n-1}^j,y_n)}(f(X))]^2, \tag{22}$$

$$M = \max_{1\leq i<j\leq N} [\mathrm{E}_{p(x_n|x_{n-1}^i,y_n)}(f(X)) - \mathrm{E}_{p(x_n|x_{n-1}^j,y_n)}(f(X))]^2. \tag{23}$$

Then

$$\frac{1}{N'}(1-\frac{1}{N_{eff}(n)})(p+\frac{m}{2}) \leq \Delta_n \leq \frac{1}{N'}(1-\frac{1}{N_{eff}(n)})(P+\frac{M}{2}). \tag{24}$$

Proof Let $S_n = \sum_{i=1}^{N} \frac{1}{N'} w_n^i (1-w_n^i)$ and $S'_n = \frac{1}{N'} \sum_{1\leq i<j\leq N} w_n^i w_n^j$. Remember that $\sum_{i=1}^{N} w_n^i = 1$ and that $1/\sum_{i=1}^{N} (w_n^i)^2 = N_{eff}(n)$. So $S_n = \frac{1}{N'}(1-\frac{1}{N_{eff}(n)})$, and

$$S'_n = \frac{1}{2N'} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} w_n^i w_n^j = \frac{1}{2N'}([\sum_{i=1}^{N} w_n^i]^2 - \sum_{i=1}^{N} (w_n^i)^2) = \frac{1}{2N'}(1-\frac{1}{N_{eff}(n)}).$$

Injecting into (16), and using (20)-(23), we get (24). □

Remark 4 From Proposition 1, let us see how Δ_n varies either in terms of $N_{eff}(n)$, for given parameters p, P, m and M, or in terms of these parameters for given $N_{eff}(n)$.

If $p \approx P$, and M (and thus m) can be neglected when compared to p, then the two bounds in (24) almost coincide and Δ_n becomes an increasing function of $N_{eff}(n)$, which means that resampling is all the more harmful as $N_{eff}(n)$ gets close to N. Observe however that for a fixed $N_{eff}(n)$, Δ_n increases as $\frac{p}{N'}$ increases; so for large values of $\frac{p}{N'}$ resampling should be avoided, even for small values of $N_{eff}(n)$.

If these conditions are not fulfilled, Δ_n is no longer stricto senso an increasing function of $N_{eff}(n)$, but remains lower- and upper- bounded by two increasing functions of $N_{eff}(n)$, and thus still tends to grow as $N_{eff}(n)$ increases. One can be more precise in two specific situations:

– If $\frac{p+\frac{m}{2}}{N'}$ is large then Δ_n gets large even for small values of $N_{eff}(n)$. Such a situation occurs in particular if, for given N' and m, p is large (in which case we turn back to the situation discussed above) or if, for given p and N', m is large, i.e. when the components of mixture $\pi_{n|n}$ are well separated.

– If $\frac{P+\frac{M}{2}}{N'}$ is small (which occurs if mixture $\pi_{n|n}$ almost reduces to a single narrow component) then Δ_n is small whatever $N_{eff}(n)$, and resampling is harmless even when $N_{eff}(n)$ is close to N.

Remark 5 From (8) and (9) Δ_n tends to 0 as N' tends to infinity, and for given parameters one can chose N' so that the upper bound of (24) becomes lower than any given threshold. So the drawbacks of resampling mentioned in Remark 4 vanish if the number of resampled particles grows to infinity. However from (8)-(9) the convergence is slow and indeed logarithmic since

$$\lim_{N'\to+\infty} \frac{|\mathrm{var}(\widehat{\Theta}^{\mathrm{SIR},N'+1}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n) - \mathrm{var}(\widehat{\Theta}^{\mathrm{SIS}}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n)|}{|\mathrm{var}(\widehat{\Theta}^{\mathrm{SIR},N'}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n) - \mathrm{var}(\widehat{\Theta}^{\mathrm{SIS}}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n)|} = 1,$$

$$\lim_{N'\to+\infty} \frac{|\mathrm{var}(\widehat{\Theta}^{\mathrm{SIR},N'+2}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n) - \mathrm{var}(\widehat{\Theta}^{\mathrm{SIR},N'+1}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n)|}{|\mathrm{var}(\widehat{\Theta}^{\mathrm{SIR},N'+1}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n) - \mathrm{var}(\widehat{\Theta}^{\mathrm{SIR},N'}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n)|} = 1.$$

4.3 SIS- vs. FA-based estimators

As we have just seen, given $\{x_{n-1}^i, w_{n-1}^i\}_{i=1}^N$ and y_n, comparing $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIR},N'})$ to $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{FA},N'})$ and $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIR},N'})$ to $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS}})$ is straightforward. It remains to compare $\widehat{\Theta}_n^{\mathrm{FA},N'}$ to $\widehat{\Theta}_n^{\mathrm{SIS}}$.

Available results [20, Section 4.2.2] [12] point out that none of these estimators systematically outperforms the other one. This indeed can be seen from (7)-(9); by identifying $\frac{N'-1}{N'}$ to 1, we see that $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{FA},N'}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n) \approx \Delta_n$; so $\widehat{\Theta}_n^{\mathrm{FA},N'}$ outperforms $\widehat{\Theta}_n^{\mathrm{SIS}}$ if, in the SIR algorithm, the extra cost Δ_n induced by resampling is lower than the original variance of $\widehat{\Theta}_n^{\mathrm{SIS}}$. In other words, whereas the final resampling step of the SIR algorithm inevitably degrades the performance of the moment estimator, resampling *before* the sampling step (i.e., running the FA algorithm) can, by contrast, be beneficial.

Let us now see more precisely which parameters play a key role in the comparison of both estimators. Let

$$\Delta'_n = \mathrm{var}(\widehat{\Theta}_n^{\mathrm{FA},N'}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n) - \mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS}}|\mathbf{x}_{n-1}^{1:N},\mathbf{w}_{n-1}^{1:N},y_n). \tag{25}$$

From (7)-(9) and (13), $\Delta'_n = \Delta_n - \frac{N'-1}{N'}\mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS}})$. By using (16) we get

$$\Delta'_n = A'_n + B_n, \tag{26}$$

in which B_n is given by (18), and

$$A'_n = \sum_{i=1}^N w_n^i(\frac{1}{N'} - w_n^i)\mathrm{var}_{p(x_n|x_{n-1}^i,y_n)}(f(X)). \tag{27}$$

Even though the sign of Δ'_n depends on function f and on the model, (26) enables some discussion. Let us begin with the following two extreme cases.

– If $N' = N$ and $w_n^i = \frac{1}{N}$ for all i, $\Delta'_n \geq 0$ and one should choose $\widehat{\Theta}_n^{\mathrm{SIS}}$. This is not surprising since in this case the resampling step of the FA algorithm would discard about one third of the particles (see the end of section 4.2.1), whereas all (equally weighted) particles are kept in the SIS algorithm.

Remark 6 Observe however that Δ'_n (which in this case reduces to B_n) decreases when the distances $|\mathrm{E}_{p(x_n|x_{n-1}^i,y_n)}(f(X)) - \mathrm{E}_{p(x_n|x_{n-1}^j,y_n)}(f(X))|$ decrease. Indeed resampling is rather harmless in this case, because after the sampling step all transformed particles $f(X_n^i)$ have almost the same mean, independently on which particles $\widetilde{x}_{n-1}^i$ had been selected (see also Remark 7 below).

– If there exists j such that $w_n^j = 1$ and $w_n^i = 0$ for all $i \neq j$ then $\Delta'_n = \frac{1-N'}{N'} \mathrm{var}_{p(x_n|x_{n-1}^j,y_n)}(f(X)) \leq 0$ and one should choose $\widehat{\Theta}_n^{\mathrm{FA},N'}$. This is not surprising either since in this extreme case, $\widehat{\Theta}_n^{\mathrm{SIS}}$ relies on only one particle ($\widehat{\Theta}_n^{\mathrm{SIS}} = f(X_n^{\mathrm{SIS},j})$ where $X_n^{\mathrm{SIS},j} \sim p(x_n|x_{n-1}^j, y_n)$), whereas $\widehat{\Theta}_n^{\mathrm{FA},N'} = \frac{1}{N'} \sum_{i=1}^{N'} f(X_n^{\mathrm{FA},i})$ in which $X_n^{\mathrm{FA},i} \sim p(x_n|x_{n-1}^j, y_n)$).

Let us now address the intermediate situations in which the weights distribution is neither uniform, nor reduced to one non null component. We have the following result.

Proposition 2 *Let p, P, m and M be defined respectively in* (20), (21), (22) *and* (23), *and let Δ'_n be defined in* (25). *Then*

$$\frac{p + \frac{m}{2}}{N'} - (P + \frac{m}{2N'}) \frac{1}{N_{eff}(n)} \leq \Delta'_n \leq \frac{P + \frac{M}{2}}{N'} - (p + \frac{M}{2N'}) \frac{1}{N_{eff}(n)}. \quad (28)$$

Proof We proceed as in the proof of Proposition 1. First, $mS'_n \leq B_n \leq MS'_n$. Next $A'_n = \frac{1}{N'} \sum_{i=1}^N w_n^i \mathrm{var}_{p(x_n|x_{n-1}^i,y_n)}(f(X)) - \sum_{i=1}^N (w_n^i)^2 \mathrm{var}_{p(x_n|x_{n-1}^i,y_n)}(f(X))$. The first term belongs to $[\frac{p}{N'}, \frac{P}{N'}]$ and the second one to $[-\frac{P}{N_{eff}(n)}, -\frac{p}{N_{eff}(n)}]$, whence (28). □

Corollary 1 *Let $N' = N$. If $1 \leq N_{eff}(n) \leq \alpha_n = \max(\frac{2pN+M}{2P+M}, 1)$ then $\widehat{\Theta}_n^{\mathrm{FA}}$ outperforms $\widehat{\Theta}_n^{\mathrm{SIS}}$. If $\beta_n = \min(\frac{2PN+m}{2p+m}, N) \leq N_{eff}(n) \leq N$ then $\widehat{\Theta}_n^{\mathrm{SIS}}$ outperforms $\widehat{\Theta}_n^{\mathrm{FA}}$.*

Proof $\widehat{\Theta}_n^{\mathrm{FA}}$ (resp. $\widehat{\Theta}_n^{\mathrm{SIS}}$) outperforms $\widehat{\Theta}_n^{\mathrm{SIS}}$ (resp. $\widehat{\Theta}_n^{\mathrm{FA}}$) if $\Delta'_n \leq 0$ (resp. ≥ 0), which is ensured if the upper (resp. lower) bound of inequality (28) ≤ 0 (resp. ≥ 0). Finally $0 < \frac{2pN+M}{2P+M} \leq N$, $\frac{2PN+M}{2p+M} \geq 1$, and remember from the discussion above that $1 \leq N_{eff}(n) \leq N$, $\Delta'_n \leq 0$ if $N_{eff}(n) = 1$ and $\Delta'_n \geq 0$ if $N_{eff}(n) = N$.

□

Remark 7 The result of corollary 1 is parametrized by p, P, m, M and N. Let us now comment some specific situations:

- For given p, P and N, $\alpha_n \to \frac{p}{P}N$ if $M \to 0$, and in particular if $p \approx P$ $\alpha_n \approx N$; in this case one should always chose $\widehat{\Theta}_n^{\mathrm{FA}}$, even when $N_{eff}(n)$ is large and close to N (this can also be seen directly from (26), in which $B_n = 0$ since $M \approx 0$ and $A'_n < 0$ since $p = P$). M very small and $p \approx P$ mean that $\mathrm{E}_{p(x_n|x_{n-1}^i, y_n)}(f(X))$ and $\mathrm{var}_{p(x_n|x_{n-1}^i, y_n)}(f(X))$ are almost independent of i, and such a situation can happen when mixture $\pi_{n|n}$ almost reduces to a single component. In this case resampling is harmless, because whichever particle $\widetilde{x}_{n-1}^i$ is selected at step i, after the sampling step the mean and variance of $f(X_n^i)$, in which $X_n^i \sim p(x_n|\widetilde{x}_{n-1}^i, y_n)$, do not depend on $\widetilde{x}_{n-1}^i$ (see also Remark 6 above).
- By contrast, for given p, P and N, $\beta_n \to 1$ if $m \to \infty$: in such a case one should choose $\widehat{\Theta}_n^{\mathrm{SIS}}$ even when $N_{eff}(n)$ is very small. Remember that in the FA algorithm $X_n^i \sim p(x_n|\widetilde{x}_{n-1}^i, y_n)$, in which $\widetilde{x}_{n-1}^i$ is the i^{th} resampled particle. Since here P is small when compared to m, points $f(X_n^i)$ and $f(X_n^j)$ are well separated if $\widetilde{x}_{n-1}^i \neq \widetilde{x}_{n-1}^j$, and in the same region otherwise. So resampling finally creates both clusters of points (all points in a given cluster are created from a particle which has been resampled several times), and void regions, because if an old particle x_{n-1}^i is discarded at the resampling step (i.e. if for all j, $\widetilde{x}_{n-1}^j \neq x_{n-1}^i$) then it is very likely that there is no compensation from the particles which have been resampled, and consequently there will be no point in the vicinity of $\mathrm{E}_{p(x_n|x_{n-1}^i, y_n)}(f(X))$.
- In practice, one can in some cases deduce values of $N_{eff}(n)$ for which $\widehat{\Theta}_n^{\mathrm{SIS}}$ outperforms $\widehat{\Theta}_n^{\mathrm{FA}}$. Consider for instance the case where $p \approx P$ and $m = 2P$. Then $\beta_n = \frac{N+1}{2}$, so for these parameters resampling is dangerous even if the number of efficient particles is weak ($\approx \frac{N}{2}$).

Remark 8 In Corollary 1 and Remark 7 we have seen that for fixed $N' = N$, the balance between the FA and SIS estimators depends on the parameters of the model and on $N_{eff}(n)$. Now $N_{eff}(n)$ is fixed for given $\mathbf{x}_{n-1}^{1:N}$, $\mathbf{w}_{n-1}^{1:N}$ and y_n, so let us see what happens at time n if we can choose $N' \neq N$. Since $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{FA},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) \to 0$ as $N' \to +\infty$, $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{FA},N'}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n) \leq \mathrm{var}(\widehat{\Theta}_n^{\mathrm{SIS}}|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n)$ if N' is large enough. Note however that depending on the parameters N' needs not be very large. From (28), $\widehat{\Theta}_n^{\mathrm{FA},N'}$ outperforms $\widehat{\Theta}_n^{\mathrm{SIS}}$ if $N' \geq \frac{1}{2p}[(2P+M)N_{eff}(n) - M]$. This lower bound tends to $\frac{P}{p}N_{eff}(n)$ when M tends to 0, i.e. when the means of the components are identical. In such a case $\widehat{\Theta}_n^{\mathrm{FA},N'}$ outperforms $\widehat{\Theta}_n^{\mathrm{SIS}}$ with less than N samples if $\frac{P}{p} \leq \frac{N}{N_{eff}(n)}$.

4.4 A hybrid SIS/FA algorithm

In the Bayesian filtering context, at a given time step one needs to propagate a discrete approximation $\widehat{\mu}_{n|n}$ of $\mu_{n|n}$ and compute an MC estimator of Θ_n (for computational concerns this second task directly makes use of the empirical measure which has just been computed). Let us for simplicity set $N' = N$. As is well known one should not compute $\widehat{\Theta}_n^{\mathrm{SIR}}$. Now, choosing at time n

between $\widehat{\Theta}_n^{\mathrm{SIS}}$ and $\widehat{\Theta}_n^{\mathrm{FA}}$ depends on $\widehat{\mu}_{n-1|n-1}$, on the new data y_n, and on the model via functions $p(x_n|x_{n-1},y_n)$ and $p(y_n|x_{n-1})$. Consequently, one can run a hybrid algorithm which automatically selects an SIS iteration or an FA one, provided of course that we know how to estimate the thresholds of Corollary 1. In summary our algorithm is as follows:

Hybrid SIS / FA algorithm.

Let $\widehat{\mu}_{n-1|n-1} = \sum_{i=1}^N w_{n-1}^i \delta_{x_{n-1}^i}$ be an MC approximation of $\mu_{n-1|n-1}$.

- For $1 \leq i \leq N$, compute $w_n^i \propto p(y_n|x_{n-1}^i) w_{n-1}^i$, $\sum_{i=1}^N w_n^i = 1$;
- Compute $N_{eff}(n)$ given in (12), and[2] α_n and β_n given in Corollary 1;
- If $N_{eff}(n) \geq \frac{\alpha_n+\beta_n}{2}$ then choose the SIS loop:
  S. For $1 \leq i \leq N$, sample x_n^i from $p(x_n|x_{n-1}^i, y_n)$;
  Then $\widehat{\mu}_{n|n} = \sum_{i=1}^N w_n^i \delta_{x_n^i}$ approximates $\mu_{n|n}$ and $\widehat{\Theta}_n^{\mathrm{SIS}} = \sum_{i=1}^N w_n^i f(x_n^i)$ approximates Θ_n.
- If $N_{eff}(n) \leq \frac{\alpha_n+\beta_n}{2}$ then choose the FA loop:
  R. For $1 \leq i \leq N$, sample $\widetilde{x}_{n-1}^i \sim \sum_{i=1}^N w_n^i \delta_{x_{n-1}^i}$;
  S. For $1 \leq i \leq N$, sample x_n^i from $p(x_n|\widetilde{x}_{n-1}^i, y_n)$.
  Then $\widehat{\mu}_{n|n} = \sum_{i=1}^N \frac{1}{N} \delta_{x_n^i}$ approximates $\mu_{n|n}$ and $\widehat{\Theta}_n^{\mathrm{FA}} = \frac{1}{N} \sum_{i=1}^N f(x_n^i)$ approximates Θ_n.

Remark 9 Note that this hybrid algorithm automatically copes with weights degeneracy: when $N_{eff}(n)$ eventually falls under the appropriate threshold then the algorithm selects the FA loop, which indeed begins with a resampling step. Now, suppose that some SIS steps have been selected up to time $n-1$, yielding $\{x_{n-1}^i, w_{n-1}^i\}_{i=1}^N$, and that the algorithm switches to the FA loop at time n. By considering that first resampling step of the n^{th} time loop as the last step of the $(n-1)^{th}$ time loop, one could think of our hybrid algorithm as being nothing but an SIR algorithm with optimal CID and optional resampling. Note however that in an SIR algorithm one would resample from $\sum_{i=1}^N w_{n-1}^i \delta_{x_{n-1}^i}$, whereas in our hybrid algorithm one resamples from $\sum_{i=1}^N w_n^i \delta_{x_{n-1}^i}$, in which $w_n^i \propto p(y_n|x_{n-1}^i) w_{n-1}^i$. So in the hybrid algorithm the resampling mechanism depends on a time varying threshold which at time n takes into account the old conditions $\{x_{n-1}^i, w_{n-1}^i\}_{i=1}^N$, but also the model and the new data y_n.

Remark 10 The hybrid algorithm is built from local (i.e., conditional) considerations, since at each time step one decides between two alternatives, depending on $\{w_n^i\}_{i=1}^N$. However the same rationale can be used to build an algorithm with unconditional properties. Compare for instance an FA-based estimator to an algorithm which propagates the particles by an FA algorithm,

[2] The bounds α_n and β_n given in Corollary 1 come from simple sufficient conditions, and no conclusion holds if $N_{eff}(n) \in (\alpha_n, \beta_n)$; so for simplicity we choose in the algorithm threshold $\frac{\alpha_n+\beta_n}{2}$. Also in practice, the time varying and difficult to compute thresholds α_n and β_n can be replaced by a common fixed value which nevertheless takes into account the parameters of the model; this will be our choice in the simulations section.

but at each n computes $\widehat{\Theta}'_n$, in which $\widehat{\Theta}'_n$ is either $\widehat{\Theta}_n^{\mathrm{SIS}}$ or $\widehat{\Theta}_n^{\mathrm{FA}}$, depending on which estimator locally outperforms the other. Then from the RB theorem $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{FA}}|\mathbf{y}_{0:n}) > \mathrm{var}(\widehat{\Theta}'_n|\mathbf{y}_{0:n})$, so $\widehat{\Theta}'_n$ unconditionally outperforms $\widehat{\Theta}_n^{\mathrm{FA}}$. But of course the price to pay is computational; by contrast with most SMC algorithms which compute the moment estimate from the set of weighted samples propagated over time, this algorithm can decide to sample new points and compute $\widehat{\Theta}_n^{\mathrm{SIS}}$, even if the points themselves are propagated via an FA loop.

5 Simulations

We now test our analysis via simulations. We also give some relative heuristics on how to adapt the threshold below which we should resample in a hybrid or SIR approach. Except in Figure 6, we take $N' = N$, i.e. the number of samples of the SMC algorithms does not vary with time.

We perform simulations in models where the FA and SIS/SIR algorithms can be implemented directly, i.e., where the optimal CID $p(x_n|x_{n-1}, y_n)$ and predictive likelihood $p(y_n|x_{n-1})$ are available (note that approximation techniques have been developed [14] [25] [26] for the cases where they cannot be computed exactly). These models include the ARCH models, widely used in financial applications, or more generally the class of Gaussian semi-linear state space models with linear observation equation but non linear state evolution equation [14]. Finally in §5.3 we adapt our analysis to conditionally linear and Gaussian jump Markov state-space systems (JMSS), for which optimal SIR and FA algorithms have also been developed [27]; the main difference is that these algorithms now act on the discrete part of the model, i.e. the Markovian jumps which monitor the state-space model. Our simulations are either local or global: by local we mean that we run a hybrid algorithm, (except in Figure 6 where we run an FA one) and at each time step n we use the common set $(\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N})$ produced by this algorithm and compute the three estimators $\widehat{\Theta}_n^{\mathrm{FA}}$, $\widehat{\Theta}_n^{\mathrm{SIS}}$ and $\widehat{\Theta}_n^{\mathrm{SIR}}$; in global simulations we run in parallel the FA, SIS, SIR and hybrid algorithms. All simulations are averaged over $L = 400$ realizations.

5.1 ARCH model

Let us consider the ARCH model:

$$\begin{cases} X_{n+1} = \sqrt{\beta_0 + \beta_1 X_n^2} \times U_n \\ Y_n = \qquad X_n + V_n \end{cases}, \tag{29}$$

in which $\{U_n\}$ and $\{V_n\}$ are i.i.d, mutually independent and independent of X_0, with $X_0 \sim \mathcal{N}(0,1)$, $U_n \sim \mathcal{N}(0,1)$, $V_n \sim \mathcal{N}(0,R)$, $\beta_0 > 0$, $\beta_1 \geq 0$. In this HMC model $p(x_n|x_{n-1}, y_n) = \mathcal{N}(\mu(x_{n-1}), \sigma^2(x_{n-1}))$ with $\mu(x_{n-1}) = \frac{\beta_0+\beta_1 x_{n-1}^2}{\beta_0+\beta_1 x_{n-1}^2+R} y_n$ and $\sigma^2(x_{n-1}) = \frac{\beta_0+\beta_1 x_{n-1}^2}{\beta_0+\beta_1 x_{n-1}^2+R} R$, and $p(y_n|x_{n-1}) = \mathcal{N}(0,\ R + \beta_0 + \beta_1 x_{n-1}^2)$. Let $N = 1000$ and $f(x) = x$. We see how to fix the threshold T_n

of the hybrid algorithm (under which we switch to the FA loop), depending on the parameters of model (29). In order to address challenging scenarios, trajectories with length $T = 60$ are generated with unknown parameters β_0^0, β_1^0 and R^0. The parameters $(\beta_0^1, \beta_1^1, R^1)$ used in our algorithms are $(1, 0.1, 9)$ for $1 \leq n \leq 20$, $(1, 1, 9)$ for $21 \leq n \leq 40$ and $(9, 5, 1)$ for $41 \leq n \leq 60$. They do not coincide with the true ones $(\beta_0^0, \beta_1^0, R^0)$, which of course will have an effect on $N_{eff}(n)$.

Let us analyse these parameters. First note that $\mu(x_{n-1})$ varies between $\frac{\beta_0}{R+\beta_0} y_n$ and y_n and $\sigma^2(x_{n-1})$ between $\frac{\beta_0}{R+\beta_0} R$ and R. Note also that only the variance of $p(y_n|x_{n-1})$ depends on x_{n-1}. In the first set of parameters β_1^1 is small, so if particles $\mathbf{x}_{n-1}^{1:N}$ are close, the means and variances of $p(x_n|x_{n-1}^i, y_n)$ will also be close, and mixture $\pi_{n|n}$ will tend to a unique component. So the FA-based estimator is preferable as soon as $N_{eff}(n)$ is below a high threshold, let us say $T_n = 0.95N$. However in this model, the dependency of $p(y_n|x_{n-1})$ in x_{n-1} is weak (again because β_1^1 is small). So $N_{eff}(n)$ will probably be close to N, and the FA- and SIS-based estimators are expected to have similar performances. In the second set $\beta_1^1 = 1$, so the components of mixture $\pi_{n|n}$ tend to become more separate and their variance will also be different. Note however that the gap between the components is bounded since $\mu(x_{n-1})$ varies between $0.1y_n$ and y_n. In this intermediate situation the threshold should be lower than the previous one, let us say $T_n = 0.8N$. Let us finally consider the informative case $\beta_0^1 = 9$, $\beta_1^1 = 5$ and $R^1 = 1$. The components of mixture $\pi_{n|n}$ get closer since the means vary between $0.9y_n$ and y_n, and the variances between $0.9R^1$ and R^1, so the FA algorithm is preferable as soon as $N_{eff}(n)$ is below a threshold higher than the two previous ones, let us say $T_n = 0.97N$.

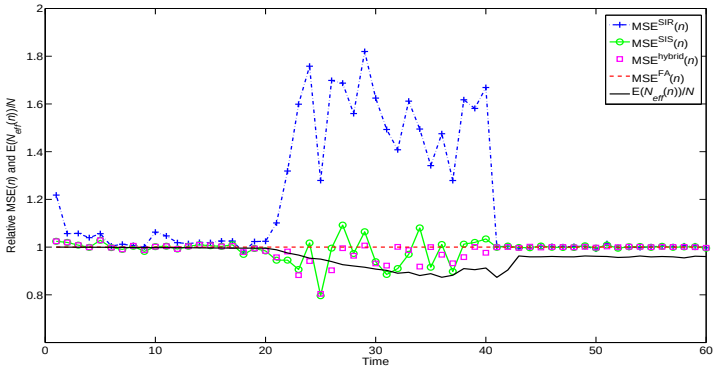

Fig. 2 Relative MSE and mean of efficient number of particles - Local Comparison - ARCH model. For $1 \leq n \leq 20$ and $41 \leq n \leq 60$, mixture $\pi_{n|n}$ tends to a unique component (more or less tight); for $21 \leq n \leq 40$, components of $\pi_{n|n}$ are more disparate.

In Figure 2, we first perform a local simulation and we plot the empirical mean square error $\mathrm{MSE}(n) = \frac{1}{L}\sum_{l=1}^{L}(\widehat{\Theta}_n(l) - \Theta_n)^2$, where $\widehat{\Theta}_n$ represents either $\widehat{\Theta}_n^{\mathrm{FA}}$, $\widehat{\Theta}_n^{\mathrm{SIS}}$ or $\widehat{\Theta}_n^{\mathrm{SIR}}$, and Θ_n defined in (3) is the true posterior mean, which here is computed by a bootstrap filter [28] with 5×10^5 particles and true parameters $(\beta_0^0, \beta_1^0, R^0)$. In the first $(1 \leq n \leq 20)$ and last $(41 \leq n \leq 60)$ parts of the experiment the SIS and FA algorithms perform similarly, since $\mathrm{E}(N_{eff}(n))$ is close to N but the components of $\pi_{n|n}$ are very close. In the second part note that $\widehat{\Theta}_n^{\mathrm{SIS}}$ often outperforms $\widehat{\Theta}_n^{\mathrm{FA}}$ even if $\mathrm{E}(N_{eff}(n))$ decreases, and that the hybrid algorithm succeeds in adapting to the best solution. Note also that the cost of resampling in the SIR algorithm depends on time. In the first part, the components of the mixture are close so resampling is harmless, while in the second part, the components are not close and resampling becomes more expensive. Finally in the last part the cost is null.

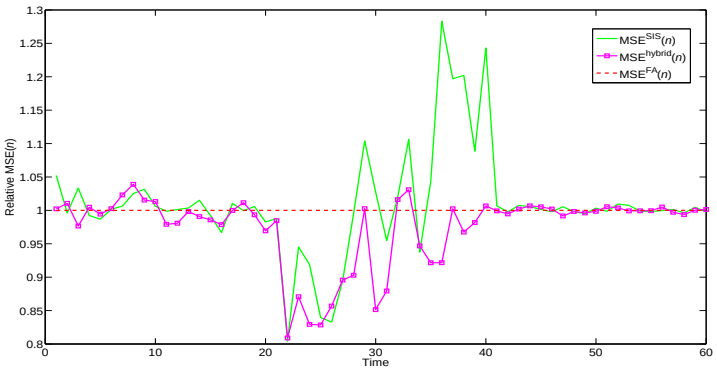

Fig. 3 Relative MSE - Global Comparison- ARCH model. In the second part the SIS threshold is too low, so $\widehat{\Theta}_n^{\mathrm{SIS}}$ degrades with time, while that of the hybrid algorithm is properly chosen.

In Figure 3 we perform a global comparison, i.e. we run independently a hybrid algorithm, an FA one, and an SIS one where we resample when $N_{eff}(n) \leq N/3$ (a value often taken in the literature). In the first and last parts, the components of the mixture are close and $\mathrm{E}(N_{eff}(n))$ is close to N. Indeed in the first part, the dependency in x_{n-1} of $p(y_n|x_{n-1})$ is weak, while in the third part, $\{x_{n-1}^i\}_{i=1}^N$ are close since they are sampled from $p(x_{n-1}|x_{n-2}^i, y_{n-1})$ which is almost independent of i and has a small variance. So when the mixture almost reduces to one component and $N_{eff}(n)$ is high all algorithms are equivalent. In the second part though, components of the mixture are still close but the model is more responsive to x_{n-1}. The threshold under which one should resample is high, which is not the case for the SIS algorithm. It is why the SIS algorithm performs poorly when compared to the hybrid algorithm at the end of the second part.

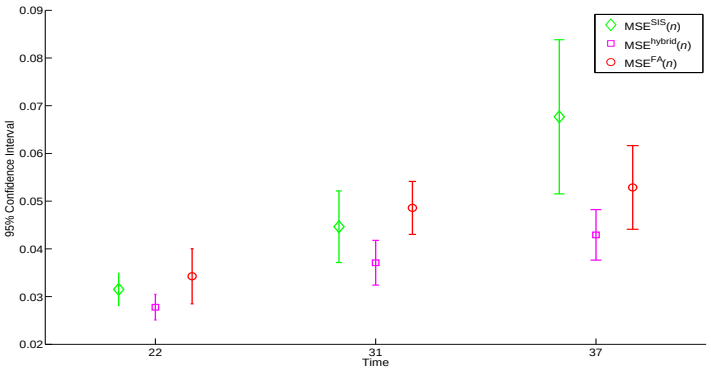

Fig. 4 95 % Confidence Intervals - Global Comparison - ARCH model. Behavior of $L = 400$ MSEs for $n = 22$, $n = 31$ and $n = 37$; $\widehat{\Theta}_n^{\mathrm{SIS}}$ and $\widehat{\Theta}_n^{\mathrm{FA}}$ are less reliable than $\widehat{\Theta}_n^{\mathrm{hybrid}}$.

Next in Figure 4, we display 95% confidence intervals built from the $L = 400$ MSEs of each algorithm. For clarity we only display three different values of the second time period $n = 22$, $n = 31$ and $n = 37$ (in the first and third periods all algorithms perform similarly), but we checked that over time the relative empirical standard deviation of the 400 MSEs has the same form as the curves in Figure 3. The SIS and FA based estimators are less reliable than the hybrid based one. Note that the SIS based estimator degrades when time increases, since its resampling threshold is not adapted to the parameters of the model. Finally we also average the RMSE over time: let $\mathcal{J} = \frac{1}{T}\sum_{i=1}^{T}\sqrt{MSE(i)}$; then $\mathcal{J}^{\mathrm{hybrid}} = 0.1173$, $\mathcal{J}^{\mathrm{SIS}} = 0.1429$ and $\mathcal{J}^{\mathrm{FA}} = 0.1305$. So it appears that minimizing the variance at each time step is a good strategy.

5.2 Gaussian model

Let us now consider a linear and Gaussian model:

$$\begin{cases} X_{n+1} = 0.9X_n + U_n \\ Y_n \quad\;\; = X_n + V_n \end{cases}, \tag{30}$$

in which U_n and V_n are i.i.d., mutually independent and independent of X_0, with $X_0 \sim \mathcal{N}(0,1)$, $U_n \sim \mathcal{N}(0,Q)$ and $V_n \sim \mathcal{N}(0,R)$. Of course Kalman filtering (KF) is available here, but our aim is this section is to illustrate different points of section 4. In this model $p(x_n|x_{n-1},y_n) = \mathcal{N}(\mu(x_{n-1}), \frac{QR}{Q+R})$ with $\mu(x_{n-1}) = 0.9x_{n-1}(1-\frac{Q}{Q+R}) + \frac{Q}{Q+R}\times y_n$ and $p(y_n|x_{n-1}) = \mathcal{N}(0.9x_{n-1}, Q+R)$.

In Figure 5 we first perform a global comparison of the FA algorithm and of two SIS ones, in which resampling is done when $N_{eff}(n) < 0.8N$ for the first one, and $0.3N$ for the second one. We take $Q = 0.01$, $R = 5$, and

$N = 1000$. For such a value of Q, pdfs $\{p(x_n|x_{n-1}^i, y_n)\}_{i=1}^N$ have very small variances (≈ 0.01) and are well separated since $1 - \frac{Q}{Q+R} \approx 1$; remember from Remark 7, item 2, that in such cases resampling is (locally) detrimental, so we expect $\widehat{\Theta}_n^{\mathrm{SIS}}$ to outperform $\widehat{\Theta}_n^{\mathrm{FA}}$, even for small values of $N_{eff}(n)$. This is confirmed by Figure 5, in which the MSE of the FA estimator is normalized to 1. By contrast to Figure 3 the gap between the SIS and FA estimators is here very large. We also notice that the SIS algorithm with threshold $0.3N$ outperforms that with threshold $0.8N$, in which resampling happens too early and eliminates essential components of the mixture (remember from Figure 1 that the number of lost components is weak when $N_{eff}(n) \approx N/3$).

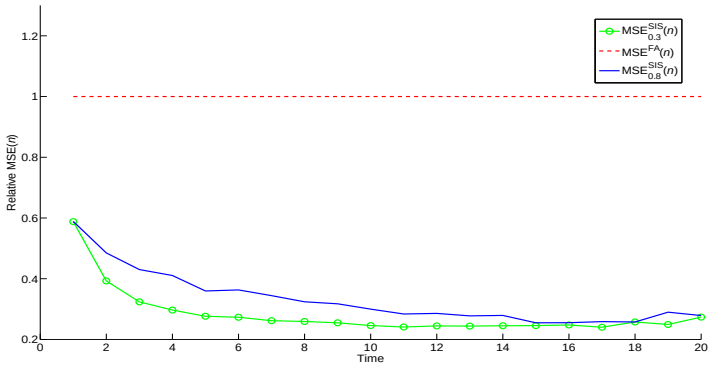

Fig. 5 Relative MSE - Global comparison - Gaussian model, $R = 5$, $Q = 0.01$. Components of mixtures $\pi_{n|n}$ are tight and separated, so resampling is detrimental.

In Figure 6 we perform a local simulation, in which we keep the same parameters, and test the effect of N'. So we compute $P(n) \stackrel{def}{=} \frac{1}{L}\sum_{l=1}^{L}(\widehat{\Theta}_n(l) - \mathrm{E}_{\pi_{n|n}}(f(X)))^2$ (an MC estimator of $\mathrm{var}(\widehat{\Theta}_n|\mathbf{x}_{n-1}^{1:N}, \mathbf{w}_{n-1}^{1:N}, y_n)$, also averaged over $L = 400$ realizations) for the FA, SIS and SIR loops, and at $n = 11$, we compare at fixed cost (see Remark 2 above) the SIR-based estimator with 2000 particles to the FA-based one with 1500 ones. Results are normalized w.r.t. $P^{\mathrm{FA}}(n)$ which is set to 1. As expected, if $N' = N$ $\widehat{\Theta}_n^{\mathrm{FA}}$ and $\widehat{\Theta}_n^{\mathrm{SIS}}$ always outperform $\widehat{\Theta}_n^{\mathrm{SIR}}$; however, since here the components of mixture $\pi_{n|n}$ have small variances and are well separated, $P^{\mathrm{SIR}}(n)$ is much larger than $P^{\mathrm{SIS}}(n)$, in accordance with Remark 4, item 1, while $\widehat{\Theta}_n^{\mathrm{FA}}$ only slightly outperforms $\widehat{\Theta}_n^{\mathrm{SIR}}$, in accordance with the end of Remark 1. For the same reason $P^{\mathrm{FA}}(n)$ is also larger than $P^{\mathrm{SIS}}(n)$, in accordance with Remark 7, item 2. Now at $n = 11$ the SIR-based estimator outperforms the FA-based one. Note that numerically $\frac{P^{\mathrm{SIS}}(11)}{P^{\mathrm{FA}}(11)} < 0.15$ and so the condition on ratio A_n holds, see Remark 2.

Let us next derive some rules of thumbs in order to tune the thresholds of the hybrid algorithm in function of the model parameters. Observe that

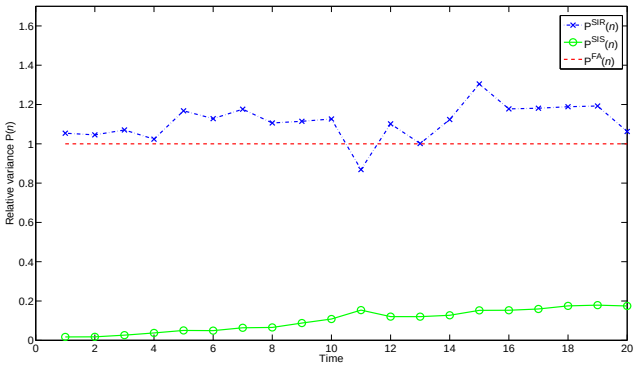

Fig. 6 Relative variance $P(n)$ - Gaussian model, $R = 5$, $Q = 0.01$. Again components of $\pi_{n|n}$ are tight and separated, so resampling is detrimental. $N = N' = 1000$, except at $n = 11$ where $N'_{\mathrm{SIR}} = 2000$ and $N'_{\mathrm{FA}} = 1500$; here $\widehat{\Theta}_{11}^{\mathrm{SIR},2000}$ outperforms $\widehat{\Theta}_{11}^{\mathrm{FA},1500}$.

in (30) Q is a key parameter, because for given $\{x_{n-1}^i\}_{i=1}^N$, j and k, the distance $|m(x_{n-1}^j) - m(x_{n-1}^k)|$ decreases with Q, while the common variance of pdfs $\{p(x_n|x_{n-1}^i, y_n)\}_{i=1}^N$ increases with Q. So if Q is small the Gaussian pdfs $\{p(x_n|x_{n-1}^i, y_n)\}_{i=1}^N$ have small variances and are well separated, and indeed we have seen in Figure 5 that $\widehat{\Theta}_n^{\mathrm{SIS}}$ outperforms $\widehat{\Theta}_n^{\mathrm{FA}}$ as long as $N_{eff}(n) > T_n$ with $T_n = 0.3N$. By contrast when Q increases the Gaussian mixture has larger and closer components; the term B_n in (18) tends to 0 and so Δ'_n in (25) tends to be negative whatever $N_{eff}(n)$ (see Remark 7, item 1). Consequently, the resampling step of the FA algorithm becomes harmless even if $N_{eff}(n)$ is high, and threshold T_n in the hybrid algorithm should be higher in this case. So we set $R = 5$, and we consider a scenario where $Q = 0.01$ for $1 \leq n \leq 20$, and we take $T_n = 0.3N$; $Q = 2$ for $21 \leq n \leq 40$, and we take $T_n = 0.65N$; $Q = 10$ for $41 \leq n \leq 60$, and we take $T_n = 0.97N$.

Figure 7 displays the results of a local simulation. The first part (i.e. $1 \leq n \leq 20$) is similar to Figure 6: since the components of $\pi_{n|n}$ are narrow and well separated it is dangerous to resample (be it before or after sampling), and indeed $\widehat{\Theta}_n^{\mathrm{SIS}}$ widely outperforms $\widehat{\Theta}_n^{\mathrm{SIR}}$, $\widehat{\Theta}_n^{\mathrm{FA}}$ outperforms $\widehat{\Theta}_n^{\mathrm{SIR}}$ only slightly, and $\widehat{\Theta}_n^{\mathrm{SIS}}$ outperforms $\widehat{\Theta}_n^{\mathrm{FA}}$ even when $N_{eff}(n)$ is weak (note however that the distance between $\widehat{\Theta}_n^{\mathrm{SIS}}$ and $\widehat{\Theta}_n^{\mathrm{FA}}$ decreases as $\mathrm{E}(N_{eff}(n))$ decreases). When Q increases, $\widehat{\Theta}_n^{\mathrm{SIS}}$ and $\widehat{\Theta}_n^{\mathrm{FA}}$ become almost equivalent for $\frac{\mathrm{E}(N_{eff}(n))}{N} \approx 0.75$ (even though $\widehat{\Theta}_n^{\mathrm{SIS}}$ remains preferable), but $\widehat{\Theta}_n^{\mathrm{FA}}$ now outperforms $\widehat{\Theta}_n^{\mathrm{SIS}}$ if $\frac{\mathrm{E}(N_{eff}(n))}{N}$ is small (see e.g. at time $n = 30$). In the third part (i.e., $41 \leq n \leq 60$) mixture $\pi_{n|n}$ almost reduces to one component; this has an influence on the SIS and FA algorithms, since $\widehat{\Theta}_n^{\mathrm{FA}}$ now outperforms $\widehat{\Theta}_n^{\mathrm{SIS}}$ (except when $\mathrm{E}(N_{eff}(n))$ is very close to N), in accordance with Remark 7, item 1; but not on the SIR algorithm, in which the resampling step remains dangerous, as expected from

Remark 1 and Remark 4, item 1. The hybrid algorithm always outperforms the others, which confirms that the thresholds are properly chosen.

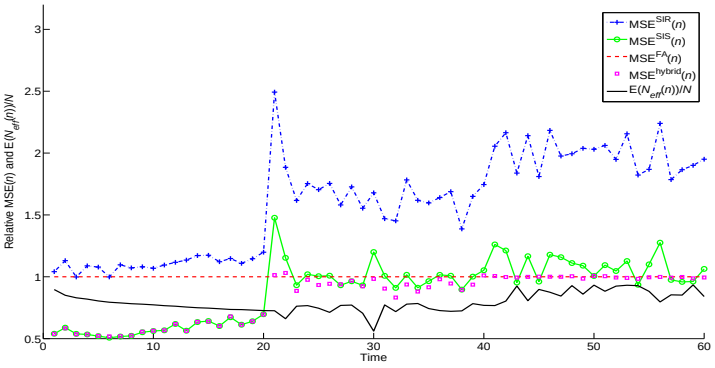

Fig. 7 Relative MSE - Local Comparison - Gaussian model. For $1 \leq n \leq 20$, components of mixture $\pi_{n|n}$ are separated and tight; for $21 \leq n \leq 40$, components are close; for $41 \leq n \leq 60$, mixture $\pi_{n|n}$ tends to a unique large component.

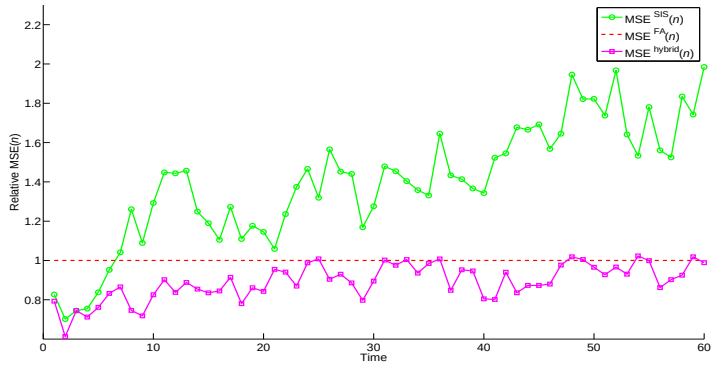

Fig. 8 Relative MSE - Global Comparison - Gaussian model. When time increases the SIS threshold is no longer adapted to the model parameters.

Finally in Figure 8 we address a challenging scenario by generating a trajectory with unknown parameters Q^0 and R^0. We perform a global comparison of an FA algorithm, an SIS one where we resample when $N_{eff}(n) \leq N/3$, and a hybrid one with threshold T_n. The parameters Q^1 and T_n/N used in our algorithms are $(1, 0.55)$ for $1 \leq n \leq 20$, $(2, 0.65)$ for $21 \leq n \leq 40$ and $(5, 0.85)$ for

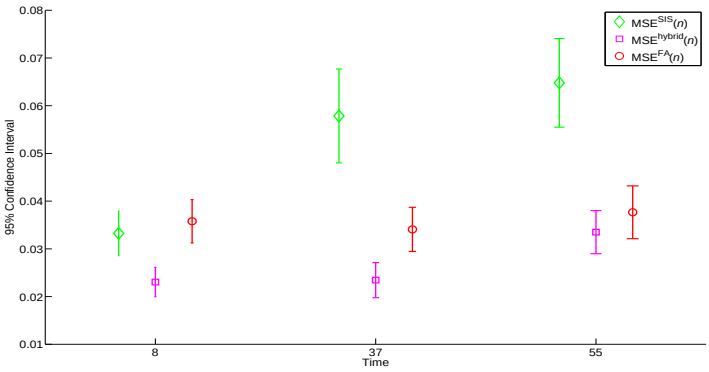

Fig. 9 95 % Confidence Intervals - Global Comparison - Gaussian model. Behavior of $L = 400$ MSEs for time steps $n = 8$, $n = 37$ and $n = 55$.

$41 \leq n \leq 60$. The MSE of the FA algorithm is normalized to 1, but is indeed rather stable over time. $\widehat{\Theta}_n^{\mathrm{SIS}}$ degrades with time, because when Q^1 increases one should indeed resample more often, and the fixed threshold $N_{eff} = N/3$ is no longer adapted to the model parameters. Also observe that the hybrid algorithm always gives the best results in this global comparison, even though it was derived from a local analysis. This is also confirmed by Figure 9 where we display 95% confidence intervals of the MSE of each algorithm for different time values, $n = 6$, $n = 37$ and $n = 55$. Again, the threshold of the SIS based algorithm is not adapted so its MSE degrades with time, while the hybrid based algorithm clearly outperforms the SIS and FA based ones. During the third time period , the performance of the FA and hybrid algorithms are close since the resampling threshold of the hybrid algorithm is actually close to 1. Finally the RMSE averaged over time are $\mathcal{J}^{\mathrm{hybrid}} = 0.1781$, $\mathcal{J}^{\mathrm{SIS}} = 0.2215$ and $\mathcal{J}^{\mathrm{FA}} = 0.1880$, which confirms the performances of the hybrid algorithm.

5.3 JMSS model

Let us finally consider a linear and Gaussian JMSS, i.e. a state-space system, the parameters of which depend on the realization of a Markov Chain $\{R_n\}_{n\geq 0}$:

$$R_n \text{ is a discrete Markov Chain,} \tag{31}$$
$$\mathbf{X}_n = \mathbf{F}_n(R_n)\mathbf{X}_{n-1} + \sigma_n(R_n)\mathbf{U}_n, \tag{32}$$
$$\mathbf{Y}_n = \mathbf{H}_n(R_n)\mathbf{X}_n + \mathbf{V}_n. \tag{33}$$

A popular algorithm in such models is the Rao-Blackwellized PF (RBPF) which consists in computing $\widehat{\Theta}_n$ via PF and KF:

$$p(\mathbf{x}_n|\mathbf{y}_{0:n}) = \underbrace{p(\mathbf{x}_n|\mathbf{r}_{0:n}, \mathbf{y}_{0:n})}_{\text{KF}} \times \underbrace{p(\mathbf{r}_{0:n}|\mathbf{y}_{0:n})}_{\text{PF}}, \tag{34}$$

$$\widehat{\Theta}_n = \sum_{i=1}^{N} w_n^i \mathrm{E}(f(\mathbf{x}_n)|\mathbf{r}_{0:n}^i, \mathbf{y}_{0:n}); \tag{35}$$

we assume that $\mathrm{E}(f(\mathbf{x}_n)|\mathbf{r}_{0:n}^i, \mathbf{y}_{0:n})$ is computable by KF. An SMC approximation $\{\mathbf{r}_{0:n}^i, w_n^i\}_{i=1}^N$ of $p(\mathbf{r}_{0:n}|\mathbf{y}_{0:n})$ is propagated by the SIS, SIR or FA procedures, except that now $w_n^i \propto w_{n-1}^i p(y_n|\mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i)$ (the second factor is computable via the prediction step of KF), and the optimal CID becomes

$$p(r_n|\mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n}) = \frac{p(\mathbf{y}_n|\mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i, r_n) p(r_n|r_{n-1}^i)}{p(\mathbf{y}_n|\mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i)}, \tag{36}$$

computable via KF too [27]. So all results of §4 are valid in such models, up to an adaptation of the notations; for instance $\mathrm{E}_{p(x_n|x_{n-1}^i, y_n)}(f(X))$ should be replaced by $\mathrm{E}_{p(r_n|\mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})}(g^i(R_n))$ where $g^i(R_n) = \mathrm{E}(f(\mathbf{X}_n)|\ \mathbf{r}_{0:n-1}^i, R_n, \mathbf{y}_{0:n})$ is computable via KF. Note that g^i depends on the whole trajectory $\mathbf{r}_{0:n-1}^i$. So means $\{\mathrm{E}_{p(r_n|\mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})}(g^i(R_n))\}_{i=1}^N$ have no reason to be close and the threshold under which we resample should be low except in some situations that we will discuss.

We deal with a target tracking scenario where $R_n \in \{1, 2, 3\}$ represents the behavior of the target (straight on, left or right turn). We have $\mathbf{X}_n = [p_x, \dot{p}_x, p_y, \dot{p}_y]_n$ and we estimate p_x. We set

$$\mathbf{F}(R) = \begin{bmatrix} 1 & \frac{\sin(\omega T_s)}{\omega} & 0 & -\frac{1-\cos(\omega T_s)}{\omega} \\ 0 & \cos(\omega T_s) & 0 & -\sin(\omega T_s) \\ 0 & \frac{1-\cos(\omega T_s)}{\omega} & 1 & \frac{\sin(\omega T_s)}{\omega} \\ 0 & \sin(\omega T_s) & 0 & \cos(\omega T_s) \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1\ 0\ 0\ 0 \\ 0\ 0\ 1\ 0 \end{bmatrix},$$

where $T_s = 2$s, $\omega = 0$ rad.s^{-1} when $R = 1$, $\omega = 5\pi/180$ rad.s^{-1} when $R = 2$ and $\omega = -5\pi/180$ rad.s^{-1} when $R = 3$,

$$\mathbf{U}_n \sim \mathcal{N}\left(0; \mathbf{Q} = \begin{bmatrix} \frac{T_s^3}{3} & \frac{T_s^2}{2} & 0 & 0 \\ \frac{T_s^2}{2} & T_s & 0 & 0 \\ 0 & 0 & \frac{T_s^3}{3} & \frac{T_s^2}{2} \\ 0 & 0 & \frac{T_s^2}{2} & T_s \end{bmatrix}\right) \text{ and } \mathbf{V}_n \sim \mathcal{N}\left(0; \mathbf{R}_v = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}\right).$$

In figures 10 and 11 we analyse the effect of the transition probabilities and of $\sigma(R)$ on the MSE of the FA, SIS and hybrid based estimators (remember that the SIR based estimator is not of practical interest), and discuss about how to fix the thresholds of the algorithms.

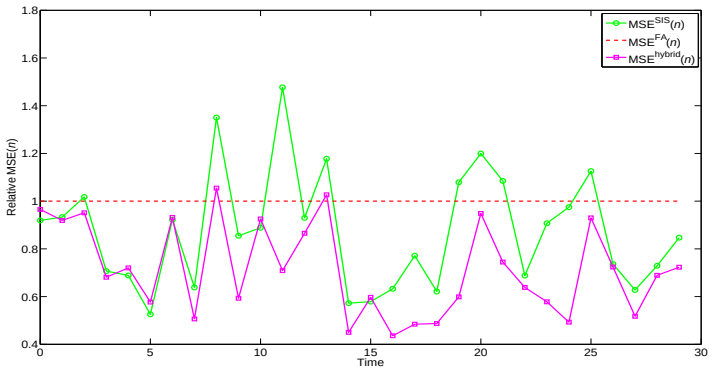

Fig. 10 Relative MSE, JMSS model. High r-dependency: $p(r_n|r_{n-1}) = 0.8$ if $r_n = r_{n-1}$ and $p(r_n|r_{n-1}) = 0.2$ if $r_n \neq r_{n-1}$; $\sigma_n(1) = 1$, $\sigma_n(2) = \sigma_n(3) = 3$. The threshold of the hybrid algorithm is $T_n = N/3$.

Let us first set $p(r_n|r_{n-1}) = 0.8$ if $r_n = r_{n-1}$, $p(r_n|r_{n-1}) = 0.1$ otherwise, and $\sigma_n(R_n) = 1$ for $R_n = 1$ and $\sigma_n(R_n) = 3$ otherwise. As pointed out above we have a diversified mixture since functions $g^i(R)$ depend on diversified trajectories $\{\mathbf{r}^i_{0:n-1}\}_{i=1}^N$. From a heuristic point of view, deleting good trajectories $\{\mathbf{r}^i_{0:n-1}\}_{i=1}^N$ during the resampling step of the FA algorithm could be dangerous because we just sample $\{r^i_n\}_{i=1}^N$ at the sampling step; in other words, it is not possible to recover lost trajectories and this can have severe consequences since g^i depend on $\mathbf{r}^i_{0:n-1}$. So the resampling step of the hybrid algorithm should only occur when $N_{eff}(n)/N \leq 0.3$. We run an FA algorithm, an SIS one where we also resample when $N_{eff}(n)/N \leq 0.3$, and the hybrid algorithm. Relative MSEs (w.r.t. the FA based estimator, which is set to 1) are plotted in Figure 10. The hybrid based estimator corresponds to the SIS based one until $k = 7$ and both outperform the FA based one. When $k \geq 7$, the hybrid based estimator outperforms the SIS based one, which means that the resampling strategy of the hybrid algorithm (which actually corresponds to an FA step, see remark 9) is more efficient than that of the SIS algorithm.

By contrast in Figure 11, we now privilege one mode by setting $p(r_n = 1|r_{n-1}) = 0.9$ and $p(r_n|r_{n-1}) = 0.05$ when $r_n \neq 1$, and reduce the dependency in R_n of the model by setting $\sigma_n(R_n) = 1$ for all R_n. So CID (36) is likely to be large when $r = 1$, we expect to have almost identical trajectories $\mathbf{r}^{1:N}_{0:n-1}$, and functions $g^i(R)$ should be close too. In this case, the threshold under which we resample should be increased, so we take $T_n = 0.75N$. We run an FA algorithm, an SIS one where we resample when $N_{eff}(n)/N \leq 0.3$, and our hybrid algorithm. The hybrid estimator corresponds to the SIS one until $k = 3$, but resampling occurs early in the hybrid algorithm and we see that it is beneficial. It seems that the threshold of the SIS algorithm is too low since after $k = 15$, the associated estimator is widely outperformed by the FA and hybrid based ones.

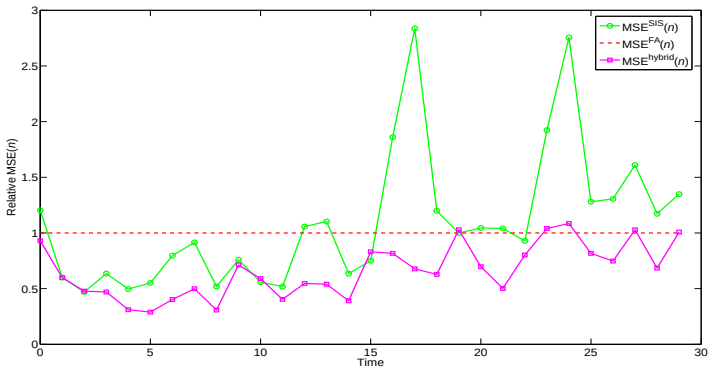

Fig. 11 Relative MSE, JMSS model. Weak r-dependency: $p(r_n = 1|r_{n-1}) = 0.9$, $p(r_n|r_{n-1}) = 0.05$ when $r_n \neq 1$. The threshold of the hybrid algorithm is $T_n = 0.75N$.

6 Conclusion

In this paper we proposed a local, non asymptotical analysis of the SIS and SIR algorithms with optimal CID, and of the FA APF algorithm; the SIR algorithm adds a resampling step as compared to the SIS one, and the SIR and FA algorithm share the same steps up to a reordering. Starting at time $n-1$ from a common empirical measure, we computed for an arbitrary finite number of particles the conditional second order moments of the three MC estimators $\widehat{\Theta}_n^{\mathrm{SIS}}$, $\widehat{\Theta}_n^{\mathrm{SIR}}$ and $\widehat{\Theta}_n^{\mathrm{FA}}$ of a moment of interest Θ_n of the filtering pdf $p_{n|n}$. We next compared $\widehat{\Theta}_n^{\mathrm{SIS}}$, $\widehat{\Theta}_n^{\mathrm{SIR}}$ and $\widehat{\Theta}_n^{\mathrm{FA}}$ two by two. First, the FA (resp. SIR) algorithm resamples before (resp. after) the sampling step; as a consequence the FA algorithm does not suffer from a lack of diversity, and $\widehat{\Theta}_n^{\mathrm{FA}}$ always outperforms $\widehat{\Theta}_n^{\mathrm{SIR}}$. Next we measured the gap between the SIS- and SIR-based estimators, in terms of the number of efficient particles $N_{eff}(n)$ and of the shape of the mixture pdf $\pi_{n|n}$ from which the points produced by the SIR algorithm are sampled from. Even though $\widehat{\Theta}_n^{\mathrm{SIS}}$ always outperforms $\widehat{\Theta}_n^{\mathrm{SIR}}$, resampling is all the more detrimental that $N_{eff}(n)$ is large, and the components of mixture $\pi_{n|n}$ get either wider or more separate. Finally, by contrast none of the SIS- and FA-based estimators systematically outperforms the other one. Nevertheless, when $\pi_{n|n}$ tends to a unique component (even of large width), $\widehat{\Theta}_n^{\mathrm{FA}}$ tends to outperform $\widehat{\Theta}_n^{\mathrm{SIS}}$ whatever $N_{eff}(n)$; if the components of $\pi_{n|n}$ get well separated $\widehat{\Theta}_n^{\mathrm{SIS}}$ outperforms $\widehat{\Theta}_n^{\mathrm{FA}}$. This analysis enabled us to propose a hybrid SIS/FA algorithm which at each instant minimizes the conditional variance of the moment estimator by selecting either the SIS or the FA loop, according to a model dependent resampling criterion. We finally validated our discussion via computer simulations.

7 Acknowledgements

The authors would like to thank the French MOD DGA/MRIS for financial support of the Ph.D. of Y. Petetin.

References

1. G. Kitagawa, Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, Journal of Computational and Graphical Statistics 5 (1) (1996) 1–25.
2. J. S. Liu, R. Chen, Sequential Monte Carlo methods for dynamic systems, Journal of the American Statistical Association 93 (443) (1998) 1032–44.
3. A. Doucet, N. de Freitas, N. Gordon (Eds.), Sequential Monte Carlo Methods in Practice, Statistics for Engineering and Information Science, Springer, New York, 2001.
4. M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear / non-Gaussian Bayesian tracking, IEEE Transactions on Signal Processing 50 (2) (2002) 174–188.
5. M. K. Pitt, N. Shephard, Filtering via simulation : Auxiliary particle filter, Journal of the American Statistical Association 94 (446) (1999) 550–99.
6. P. Fearnhead, Computational methods for complex stochastic systems: A review of some alternatives to MCMC, Statistics and Computing 18 (2) (2008) 151–71.
7. N. Whiteley, A. M. Johansen, Recent developments in auxiliary particle filtering, in: Barber, Cemgil, Chiappa (Eds.), Inference and Learning in Dynamic Models, Cambridge University Press, 2010.
8. H. R. Künsch, State space and hidden Markov models, in: O. E. Barndorff-Nielsen, D. R. Cox, C. Klüppelberg (Eds.), Complex Stochastic Systems, Vol. 87 of Monographs on Statistics and Applied Probability, Chapman and Hall / CRC, New York, 2001, pp. 109–73.
9. D. Crisan, A. Doucet, A survey of convergence results on particle filtering methods for practitioners, IEEE Transactions on Signal Processing 50 (3) (2002) 736–46.
10. P. del Moral, Feynman-Kac formulae. Genealogical and interacting particle systems with applications, Probability and its applications, Springer, New-York, 2004.
11. O. Cappé, É. Moulines, T. Rydén, Inference in Hidden Markov Models, Springer, 2005.
12. A. M. Johansen, A. Doucet, A note on the auxiliary particle filter, Statistics and Probability Letters 78 (12) (2008) 1498–1504.
13. R. Douc, É. Moulines, J. Olsson, Optimality of the auxiliary particle filter, Probability and Mathematical Statistics 29 (1) (2009) 1–28.
14. A. Doucet, S. J. Godsill, C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, Statistics and Computing 10 (2000) 197–208.
15. H. Künsch, Recursive Monte Carlo filters: Algorithms and theoretical analysis, The Annals of Statistics 33 (5) (2005) 1983–2021.
16. R. Douc, O. Cappé, É. Moulines, Comparison of resampling schemes for particle filtering, in: Proc. of the 4th ISPA, Zagreb, Croatia, 2005.
17. J. D. Hol, T. B. Schön, F. Gustafsson, On resampling algorithms for particle filtering, in: Proc. IEEE NSSPW, Cambridge, UK, 2006.
18. V. Zaritskii, V. Svetnik, L. Shimelevich, Monte Carlo technique in problems of optimal data processing, Automation and remote control (1975) 95–103.
19. A. Kong, J. S. Liu, W. H. Wong, Sequential imputations and bayesian missing data problems, Journal of the American Statistical Association 89 (425) (1994) 278–88.
20. J. S. Liu, R. Chen, Blind deconvolution via sequential imputation, Journal of the American Statistical Association 90 (430) (1995) 567–76.
21. B. D. Ripley, Stochastic simulation, Wiley, 1987.
22. F. Desbouvries, Y. Petetin, E. Monfrini, Optimal SIR algorithm vs. fully adapted auxiliary particle filter: a matter of conditional independence, in: Proc. IEEE ICASSP, Prague, Czech Republic, 2011.
23. J. S. Liu, Metropolized independent sampling with comparisons to rejection sampling and importance sampling, Statistics and Computing 6 (1996) 113–119.

24. J. Cornebise, É. Moulines, J. Olsson, Adaptive methods for sequential importance sampling with application to state-space models, Statistics and Computing 18 (4) (2008) 461–480.
25. S. Saha, P. K. Manda, Y. Boers, H. Driessen, A. Bagchi, Gaussian proposal density using moment matching in SMC methods., Statistics and Computing 19 (2) (2009) 203–208.
26. R. van der Merwe, A. Doucet, N. de Freitas, E. Wan, The unscented particle filter, Advances in Neural Information Processing Systems.
27. A. Doucet, N. J. Gordon, V. Krishnamurthy, Particle filters for state estimation of jump Markov linear systems, IEEE Transactions on Signal Processing 49 (3) (2001) 613–24.
28. N. J. Gordon, D. J. Salmond, A. F. M. Smith, Novel approach to nonlinear/ non-Gaussian Bayesian state estimation, IEE Proceedings-F 140 (2) (1993) 107–113.