A NON ASYMPTOTICAL ANALYSIS OF THE OPTIMAL SIR ALGORITHM VS. THE FULLY ADAPTED AUXILIARY PARTICLE FILTER

François Desbouvries, Yohan Petetin and Emmanuel Monfrini

Telecom Institute / Telecom SudParis / CITI Department and CNRS UMR 5157,
9 rue Charles Fourier, 91011 Evry, France

ABSTRACT

Particle filters (PF) and auxiliary particle filters (APF) are widely used sequential Monte Carlo (SMC) techniques for estimating the a posteriori filtering pdf in a Hidden Markov Chain (HMC). These algorithms have been theoretically analysed from an asymptotical statistics perspective. In this paper we provide a *non asymptotical, finite number of samples* comparative analysis of two particular SMC algorithms : the Sampling Importance Resampling (SIR) PF with optimal conditional importance distribution (CID), and the fully adapted APF (FA). Starting from a common set of N particles, we compute closed form expressions of the mean and variance of the empirical Monte Carlo (MC) estimators of a moment of the a posteriori filtering pdf. Both algorithms have the same mean, but in the case where resampling is used, the variance of the SIR algorithm always exceeds that of the FA algorithm.

Index Terms— Sequential Monte Carlo, Auxiliary Particle Filtering, non asymptotical analysis

1. INTRODUCTION

Let X_n and Y_n be respectively a hidden and an observed process. Let $p(x_n|\mathbf{y}_{0:n})$ (or in short $p_{n|n}$) denote the a posteriori filtering pdf of x_n given $\mathbf{y}_{0:n} = \{y_i\}_{i=0}^n$. We assume that $\{X_n, Y_n\}_{n \in \mathbb{N}}$ is an HMC : $p(\mathbf{x}_{0:n}, \mathbf{y}_{0:n}) = p(x_0) \prod_{i=1}^n p(x_i|x_{i-1}) \prod_{i=0}^n p(y_i|x_i)$. Bayesian filtering consists in computing $p(x_n|\mathbf{y}_{0:n})$, or at least some approximation of the measure $p(dx_n|\mathbf{y}_{0:n})$ with density $p(x_n|\mathbf{y}_{0:n})$.

Recursive solutions are of particular interest, and indeed $p_{n|n}$ can be computed from $p_{n-1|n-1}$ by the well known recursion (here $\mathcal{N}$ stands for numerator) :

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

Many efforts have been devoted to the computation of eq. (1). If exact computing is difficult or impossible, one needs to resort to approximations. Among them PF and APF methods

(see e.g. [1]-[4]) are SMC methods which propagate a discrete approximation $\hat{p}(dx_n|\mathbf{y}_{0:n})$ of $p(dx_n|\mathbf{y}_{0:n})$.

From a theoretical point of view, convergence results of some SMC algorithms have been proven in a series of papers, see e.g. [5]-[8] and references therein. These results have been recently extended to the APF [9], and a tuning of the first-stage weights of the APF from an asymptotical variance analysis viewpoint has also been proposed in [10].

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 one particular SMC algorithm : the SIR algorithm with so-called optimal CID [11], versus the FA [3]. From a technical point of view, such a finite samples analysis is indeed feasible because in the case of this optimal SIR algorithm the incremental weight does not depend on the particle which has just been sampled, which enables closed form mean and variance computations.

From this finite samples analysis we derive the following conclusion. Assume that we aim to compute some moment

$$\Theta_n = \mathrm{E}_{p_{n|n}}(f(X)) \tag{2}$$

(here $f(.)$ is some function of interest) of the true pdf $p_{n|n}$. Starting from a common set of points (assumed to be at least approximately sampled from $p_{n-1|n-1}$), the MC empirical estimates $\hat{\Theta}_n^{\mathrm{SIR}}$ and $\hat{\Theta}_n^{\mathrm{FA}}$ of Θ_n computed respectively by the SIR and FA algorithms have the same conditional mean, but the conditional variance of $\hat{\Theta}_n^{\mathrm{SIR}}$ always exceeds that of $\hat{\Theta}_n^{\mathrm{FA}}$. This paper is organized as follows. The SIR and FA algorithms are recalled in §2. Our analysis is performed in §3, and we discuss the results in section §4.

2. TWO CLOSELY RELATED SMC FILTERS

Let us first recall the well known SIR algorithm with optimal CID (see e.g. [11]). It consists of the succession of a sampling step *S*, a weighting step *W* and an optional resampling step *R*. The algorithm (which will simply be denoted as SIR in the sequel) is as follows :

SIR algorithm (with optimal CID and Resampling at each step).

Let $\widehat{p}(dx_{n-1}|\mathbf{y}_{0:n-1}) = \sum_{i=1}^{N} \frac{1}{N} \delta_{x_{n-1}^{\mathrm{SIR},i}}(dx_{n-1})$.

S. For $1 \le i \le N$, sample $x_n^{\mathrm{IS},i}$ from $p(x_n|x_{n-1}^{\mathrm{SIR},i}, y_n)$;

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

R. For $1 \le i \le N$, sample $x_n^{\mathrm{SIR},i} \sim \sum_{i=1}^{N} w_n^i \delta_{x_n^{\mathrm{IS},i}}(dx_n)$.

Then $\widehat{p}(dx_n|\mathbf{y}_{0:n}) = \sum_{i=1}^{N} \frac{1}{N} \delta_{x_n^{\mathrm{SIR},i}}(dx_n) \simeq p(dx_n|\mathbf{y}_{0:n})$.

Let us now briefly recall the principle of APF [3] [4] (see also [12] for recent developments). If exact computing of (1) is not possible, a natural idea consists in replacing $p(dx_{n-1}|\mathbf{y}_{0:n-1})$ by a discrete approximation $\sum_{i=1}^{N} w_{n-1}^i \delta_{x_{n-1}^i}(dx_{n-1})$. Injecting into (1) we get a continuous approximation of $p(x_n|\mathbf{y}_{0:n})$:

$$\pi_n(x_n) = \sum_{i=1}^{N} \frac{w_{n-1}^i p(y_n|x_{n-1}^i)}{\sum_{i=1}^{N} w_{n-1}^i p(y_n|x_{n-1}^i)} p(x_n|x_{n-1}^i, y_n). \quad (3)$$

In the view of deriving an SMC algorithm one should get samples from (3) in order to further proceed at time $n+1$. If it is not possible to sample from π_n one can use importance sampling and we then get APF. Now, if direct sampling from π_n is indeed feasible then we are in the so-called *fully-adapted* (FA) case, in the terminology of [3]. Remember that to get a sample from a mixture $\sum_{i=1}^{N} \alpha_i p_i(x)$ one can first draw an index j from $\sum_{i=1}^{N} \alpha_i \delta_i$ ((W,R) steps) and then sample x from the $p_j(x)$ which has just been selected (S step). Applying to (3) we get the

FA algorithm[1]

Let $\widehat{p}(dx_{n-1}|\mathbf{y}_{0:n-1}) = \sum_{i=1}^{N} \frac{1}{N} \delta_{x_{n-1}^{\mathrm{FA},i}}(dx_{n-1})$.

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

R. For $1 \le i \le N$, sample $\tilde{x}_{n-1}^i \sim \sum_{i=1}^{N} w_n^i \delta_{x_{n-1}^{\mathrm{FA},i}}(dx_{n-1})$;

S. For $1 \le i \le N$, sample $x_n^{\mathrm{FA},i}$ from $p(x_n|\tilde{x}_{n-1}^i, y_n)$.

Then $\widehat{p}(dx_n|\mathbf{y}_{0:n}) = \sum_{i=1}^{N} \frac{1}{N} \delta_{x_n^{\mathrm{FA},i}}(dx_n) \simeq p(dx_n|\mathbf{y}_{0:n})$.

As we see the FA algorithm is a *reordering* of the SIR algorithm. In SIR the successive steps are $S \to W \to R$ while the recursive loop of FA is made of the successive steps $W \to R \to S$.

3. A NON ASYMPTOTICAL ANALYSIS OF THE SIR AND FA ALGORITHMS

3.1. Reformulating the SIR and FA algorithms

In view of §3.2 we first need to rewrite the algorithms in terms of conditional pdf.

---

[1]In the view of §3 we set $w_{n-1}^i = \frac{1}{N}$ for all i, i.e. we start from an unweighted pdf.

- In FA, given $\mathbf{x}_{n-1}^{\mathrm{FA},\{i\}} \stackrel{def}{=} \{x_{n-1}^{\mathrm{FA},i}\}_{i=1}^{N}$ and given y_n, we draw N independent points $\mathbf{x}_n^{\mathrm{FA},\{i\}}$ from $\pi_n(.)$ defined in (3).

- Let us now consider the SIR algorithm. Let $\overline{p}(y_n|x_{n-1}^i) = \frac{p(y_n|x_{n-1}^i)}{\sum_{i=1}^{N} p(y_n|x_{n-1}^i)}$. In the (S) step, given $\mathbf{x}_{n-1}^{\mathrm{SIR},\{i\}} \stackrel{def}{=} \{x_{n-1}^{\mathrm{SIR},i}\}_{i=1}^{N}$ and given y_n, we independently draw N particles $\mathbf{x}_n^{\mathrm{IS},\{i\}} \stackrel{def}{=} \{x_n^{\mathrm{IS},i}\}_{i=1}^{N}$. We next weight them, and given $\mathbf{x}_{n-1}^{\mathrm{SIR},\{i\}}$ and $\mathbf{x}_n^{\mathrm{IS},\{i\}}$, we draw N independent samples $\mathbf{x}_n^{\mathrm{SIR},\{i\}}$ from this reweighted distribution (R step). This algorithm can be rewritten as follows :

SIR algorithm. Let $\mathbf{x}_{n-1}^{\mathrm{SIR},\{i\}} \stackrel{i.i.d.}{\sim} p(x_{n-1}|\mathbf{y}_{0:n-1})$. Then

$$\mathbf{x}_n^{\mathrm{IS},\{i\}} \stackrel{S}{\sim} p(\mathbf{x}_n|\mathbf{x}_{n-1}^{\mathrm{SIR},\{i\}}, y_n)$$

$$= \prod_{i=1}^{N} p(x_n^i|x_{n-1}^{\mathrm{SIR},i}, y_n); \quad (4)$$

$$\mathbf{x}_n^{\mathrm{SIR},\{i\}} \stackrel{(W,R)}{\sim} p(\mathbf{x}_n|\mathbf{x}_n^{\mathrm{IS},\{i\}}, \mathbf{x}_{n-1}^{\mathrm{SIR},\{i\}}, y_n)$$

$$= \prod_{i=1}^{N} (\sum_{j=1}^{N} \overline{p}(y_n|x_{n-1}^{\mathrm{SIR},j}) \delta_{x_n^{\mathrm{IS},j}}(x_n^i)). \quad (5)$$

3.2. Second order moments of the MC estimators

Let us now assume that our aim is to compute the moment Θ_n defined in (2). So assume that at time $n-1$ we have at our disposal a set of N points $\mathbf{x}_{n-1}^{\{i\}}$ (approximately) sampled from $p_{n-1|n-1}$. Starting from this common set of points, the FA (resp. SIR) algorithm produces a new set of samples $\mathbf{X}_n^{\mathrm{FA},\{i\}}$ (resp. $\mathbf{X}_n^{\mathrm{IS},\{i\}}$ and $\mathbf{X}_n^{\mathrm{SIR},\{i\}}$) which in turn can be used to compute the following estimators of Θ_n :

$$\hat{\Theta}_n^{\mathrm{FA}}(\mathbf{X}_n^{\mathrm{FA},\{i\}}) = \frac{1}{N} \sum_{i=1}^{N} f(X_n^{\mathrm{FA},i}), \quad (6)$$

$$\hat{\Theta}_n^{\mathrm{IS}}(\mathbf{X}_n^{\mathrm{IS},\{i\}}) = \sum_{i=1}^{N} \overline{p}(y_n|x_{n-1}^i) f(X_n^{\mathrm{IS},i}), \quad (7)$$

$$\hat{\Theta}_n^{\mathrm{SIR}}(\mathbf{X}_n^{\mathrm{SIR},\{i\}}) = \frac{1}{N} \sum_{i=1}^{N} f(X_n^{\mathrm{SIR},i}); \quad (8)$$

So in the case of the SIR algorithm $\hat{\Theta}_n^{\mathrm{IS}}$ (resp. $\hat{\Theta}_n^{\mathrm{SIR}}$) is the estimator built before (resp. after) the resampling step. Our aim in this section is to compare, *for a finite number of samples N*, the mean and variances of estimators (6), (7) and (8).

Theorem 1 Let $\hat{\Theta}_n^{\mathrm{FA}}$, $\hat{\Theta}_n^{\mathrm{IS}}$ and $\hat{\Theta}_n^{\mathrm{SIR}}$ be defined as in (6)-(8), and $\pi_n(.)$ be given by (3). Then given $\mathbf{x}_{n-1}^{\{i\}}$ and y_n,

$$\mathrm{E}(\hat{\Theta}_n^{\mathrm{FA}}) = \mathrm{E}_{\pi_n}(f(X)), \quad (9)$$
$$\mathrm{E}(\hat{\Theta}_n^{\mathrm{IS}}) = \mathrm{E}_{\pi_n}(f(X)), \quad (10)$$
$$\mathrm{E}(\hat{\Theta}_n^{\mathrm{SIR}}) = \mathrm{E}_{\pi_n}(f(X)). \quad (11)$$

Proof 1 *Given* $\mathbf{x}_{n-1}^{\{i\}}$ *and* y_n, $\mathbf{X}_n^{FA,i}$ *are sampled from* $\pi_n(.)$ *by construction, whence (9). Next, given* $\mathbf{x}_{n-1}^{\{i\}}$ *and* y_n,

$$\mathrm{E}(\hat{\Theta}_n^{\mathrm{IS}}) \stackrel{(4)}{=} \sum_{i=1}^N \overline{p}(y_n|x_{n-1}^i) \int_{\mathrm{I\!R}} f(x_n^i) p(x_n^i|x_{n-1}^i, y_n) dx_n^i$$

$$= \int_{\mathrm{I\!R}} f(x) \underbrace{\sum_{i=1}^N \overline{p}(y_n|x_{n-1}^i) p(x|x_{n-1}^i, y_n)}_{\pi_n(x)} dx$$

$= \mathrm{E}_{\pi_n}(f(X))$, *whence (10). Now from (5), for all* i,

$$\mathrm{E}(f(X_n^{\mathrm{SIR},i})|\mathbf{x}_n^{\mathrm{IS},\{i\}}) = \hat{\Theta}_n^{\mathrm{IS}}(\mathbf{x}_n^{\mathrm{IS},\{i\}}). \qquad (12)$$

So, given $\mathbf{x}_{n-1}^{\{i\}}$ *and* y_n,

$$\begin{aligned}
\mathrm{E}(f(X_n^{\mathrm{SIR},i})) &= \mathrm{E}(\mathrm{E}(f(X_n^{\mathrm{SIR},i})|\mathbf{X}_n^{\mathrm{IS},\{i\}})) \\
&\stackrel{(12)}{=} \mathrm{E}(\hat{\Theta}_n^{\mathrm{IS}}) & (13) \\
&\stackrel{(10)}{=} \mathrm{E}_{\pi_n}(f(X)), & (14)
\end{aligned}$$

whence (11). ∎

Theorem 2 *Let* $\hat{\Theta}_n^{FA}$, $\hat{\Theta}_n^{IS}$ *and* $\hat{\Theta}_n^{SIR}$ *be defined as in (6)-(8), and* $\pi_n(.)$ *be given by (3). Then given* $\mathbf{x}_{n-1}^{\{i\}}$ *and* y_n,

$$\mathrm{var}(\hat{\Theta}_n^{\mathrm{FA}}) = \frac{1}{N}\mathrm{var}_{\pi_n}(f(X)), \qquad (15)$$

$$\mathrm{var}(\hat{\Theta}_n^{\mathrm{IS}}) = \sum_{i=1}^N (\overline{p}(y_n|x_{n-1}^i))^2 \mathrm{var}_{p(.|x_{n-1}^i,y_n)}(f(X)), \ (16)$$

$$\mathrm{var}(\hat{\Theta}_n^{\mathrm{SIR}}) = \frac{1}{N}\mathrm{var}_{\pi_n}(f(X)) \qquad (17)$$
$$+ \frac{N-1}{N}\sum_{i=1}^N (\overline{p}(y_n|x_{n-1}^i))^2 \mathrm{var}_{p(.|x_{n-1}^i,y_n)}(f(X)).$$

Proof 2 *Given* $\mathbf{x}_{n-1}^{\{i\}}$ *and* y_n, $\mathbf{X}_n^{FA,\{i\}}$ *are i.i.d. from* $\pi_n(.)$ *by construction, whence (15); and similarly from (4)* $\mathbf{X}_n^{IS,\{i\}}$ *are sampled independently, and each point* $X_n^{IS,i}$ *is sampled from* $p(.|\mathbf{x}_{n-1}^{\{i\}}, y_n)$, *whence (16). Given* $\mathbf{x}_{n-1}^{\{i\}}$ *and* y_n, *let us now compute*

$$\begin{aligned}
\mathrm{var}(\hat{\Theta}_n^{\mathrm{SIR}}) &= \frac{1}{N^2}\sum_{i=1}^N \mathrm{var}(f(X_n^{\mathrm{SIR},i})) \\
&+ \frac{1}{N^2}\sum_{\substack{k,l=1 \\ k\neq l}}^N \mathrm{Cov}(f(X_n^{\mathrm{SIR},k}), f(X_n^{\mathrm{SIR},l})). (18)
\end{aligned}$$

Let us begin with the first term. Let $1 \leq i \leq N$. *From (5),* $\mathrm{E}(f^2(X_n^{\mathrm{SIR},i})|\mathbf{x}_n^{\mathrm{IS},\{i\}}) = \sum_{i=1}^N \overline{p}(y_n|x_{n-1}^i)f^2(x_n^{\mathrm{IS},i})$; *so given* $\mathbf{x}_{n-1}^{\{i\}}$ *and* y_n,

$$\mathrm{E}(f^2(X_n^{\mathrm{SIR},i})) = \mathrm{E}(\mathrm{E}(f^2(X_n^{\mathrm{SIR},i})|\mathbf{X}_n^{IS,\{i\}}))$$

$$\stackrel{(4)}{=} \int_{\mathrm{I\!R}^N} (\sum_{i=1}^N \overline{p}(y_n|x_{n-1}^i)f^2(x_n^i)) \prod_{j=1}^N p(x_n^j|x_{n-1}^j, y_n) d\mathbf{x}_n^{\{i\}}$$

$= \mathrm{E}_{\pi_n}(f^2(X))$, *and using (14) we see that the first term of (18) is equal to* $\frac{1}{N}\mathrm{var}_{\pi_n}(f(X))$. *Let us now compute the second term. Using (5) and (12), we see that for all* k, l, $1 \leq k, l, \leq N$ *with* $k \neq l$,

$$\mathrm{E}(f(X_n^{\mathrm{SIR},k})f(X_n^{\mathrm{SIR},l})|\mathbf{x}_n^{\mathrm{IS},\{i\}}) = (\hat{\Theta}_n^{\mathrm{IS}}(\mathbf{x}_n^{\mathrm{IS},\{i\}}))^2.$$

So given $\mathbf{x}_{n-1}^{\{i\}}$ *and* y_n, $\mathrm{E}(f(X_n^{\mathrm{SIR},k})f(X_n^{\mathrm{SIR},l})) =$

$$\mathrm{E}(\mathrm{E}(f(X_n^{\mathrm{SIR},k})f(X_n^{\mathrm{SIR},l})|\mathbf{X}_n^{\mathrm{IS},\{i\}}))$$

$= \mathrm{E}((\hat{\Theta}_n^{\mathrm{IS}})^2)$. *Using (13), we conclude that given* $\mathbf{x}_{n-1}^{\{i\}}$ *and* y_n,

$$\mathrm{Cov}(f(X_n^{\mathrm{SIR},k}), f(X_n^{\mathrm{SIR},l})) = \mathrm{var}(\hat{\Theta}_n^{\mathrm{IS}}). \qquad (19)$$

Injecting into (18) we finally get (17). ∎

4. DISCUSSION

From Theorem 1, given $\mathbf{x}_{n-1}^{\{i\}}$ and y_n, the three estimators $\hat{\Theta}_n^{FA}$, $\hat{\Theta}_n^{IS}$ and $\hat{\Theta}_n^{SIR}$ have the same mean. So we compare them from their conditional variances only.

4.1. SIR vs. FA algorithms.

From (15) and (17) we see that for all N,

$$\mathrm{var}(\hat{\Theta}_n^{\mathrm{SIR}}|\mathbf{x}_{n-1}^{\{i\}}, y_n) > \mathrm{var}(\hat{\Theta}_n^{\mathrm{FA}}|\mathbf{x}_{n-1}^{\{i\}}, y_n).$$

Starting from a common set of points, it is thus preferable to perform the *(W,R,S)* operations in that order, rather than sampling first, then weighting and resampling.

4.2. SIR vs. IS algorithms.

Even though (17) decomposes $\mathrm{var}(\hat{\Theta}_n^{SIR}|\mathbf{x}_{n-1}^{\{i\}}, y_n)$ as the sum of two positive terms, it should not be confused with the Rao-Blackwell equality which in the context of this paper reads

$$\begin{aligned}
\mathrm{var}(\hat{\Theta}_n^{\mathrm{SIR}}|\mathbf{x}_{n-1}^{\{i\}}, y_n) &= \mathrm{var}(\hat{\Theta}_n^{\mathrm{IS}}|\mathbf{x}_{n-1}^{\{i\}}, y_n) & (20) \\
&+ \mathrm{E}((\hat{\Theta}_n^{\mathrm{SIR}} - \hat{\Theta}_n^{\mathrm{IS}})^2|\mathbf{x}_{n-1}^{\{i\}}, y_n).
\end{aligned}$$

Equation (20) enables to conclude, as is well known [8, p. 213], that

$$\mathrm{var}(\hat{\Theta}_n^{\mathrm{SIR}}|\mathbf{x}_{n-1}^{\{i\}}, y_n) > \mathrm{var}(\hat{\Theta}_n^{\mathrm{IS}}|\mathbf{x}_{n-1}^{\{i\}}, y_n).$$

So it is preferable to compute $\hat{\Theta}_n$ from the weighted set of points $\{(x_n^{\mathrm{IS},\{i\}}, \overline{p}(y_n|x_{n-1}^i)\}_{i=1}^N$ rather than from the unweighted one $\{(x_n^{\mathrm{SIR},\{i\}}, 1/N)\}_{i=1}^N$; the second term in (20) is the price we pay when resampling the particles.

Now, notice that in the context of this paper (20) can be computed : $\mathrm{var}(\hat{\Theta}_n^{\mathrm{IS}}|\mathbf{x}_{n-1}^{\{i\}}, y_n)$ is given by (16), and

$$E((\hat{\Theta}_n^{\mathrm{SIR}} - \hat{\Theta}_n^{\mathrm{IS}})^2|\mathbf{x}_{n-1}^{\{i\}}, y_n) = \frac{1}{N}\mathrm{var}_{\pi_n}(f(X))$$

$$-\frac{1}{N}\sum_{i=1}^{N}(\overline{p}(y_n|x_{n-1}^i))^2\mathrm{var}_{p(.|x_{n-1}^i,y_n)}(f(X)).$$

4.3. IS vs. FA algorithms.

As we have just seen, given $\mathbf{x}_{n-1}^{\{i\}}$ and y_n comparing $\mathrm{var}(\hat{\Theta}_n^{\mathrm{SIR}}$, to $\mathrm{var}(\hat{\Theta}_n^{\mathrm{FA}})$ and $\mathrm{var}(\hat{\Theta}_n^{\mathrm{SIR}})$ to $\mathrm{var}(\hat{\Theta}_n^{\mathrm{IS}})$ is straightforward. Interestingly enough, even if we have a closed form espression of

$$\Delta_n = \mathrm{var}(\hat{\Theta}_n^{\mathrm{IS}}) - \mathrm{var}(\hat{\Theta}_n^{\mathrm{FA}})$$

(which describes the behavior of the SIR algorithm *before resampling* as compared to the FA algorithm), Δ_n can be computed explicitly for some functions f and HMC models only, and thus its sign is hard to predict in general. Finally as we shall see from Fig. 1, Δ_n can be positive or negative, which somehow confirms the conclusions (drawn from an asymptotical analysis viewpoint) of [9].

4.4. Simulations

Let us finally perform a simulation. Let us consider the linear and Gaussian model :

$$\begin{cases} x_{n+1} &= 0.2x_n + u_n \\ y_n &= 5x_n + v_n \end{cases}, \tag{21}$$

in which u_n and v_n are i.i.d., mutually independent and independent of x_0, with $x_0 \sim \mathcal{N}(0,1)$. Let also $u_n \sim \mathcal{N}(0,Q)$, $Q = 10$ and $v_n \sim \mathcal{N}(0,R)$, $R = 1$. We run the FA algorithm with $N = 1000$ particles; at each time step, we use the common set $\{x_{n-1}^i\}_{i=1}^N$ produced by this algorithm and compute the three estimators. In model (21) the true mean Θ_n can be computed by the Kalman Filter, so for each algorithm we plot the empirical MSE (averaged over $P = 1000$ realizations). In Fig. 1 these empirical MSE are normalized w.r.t. that of the FA algorithm, which is set to 1.

5. REFERENCES

[1] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer Verlag, New York, 2001.

[2] M. Sanjeev Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear / non-Gaussian Bayesian tracking," *IEEE Tr. Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
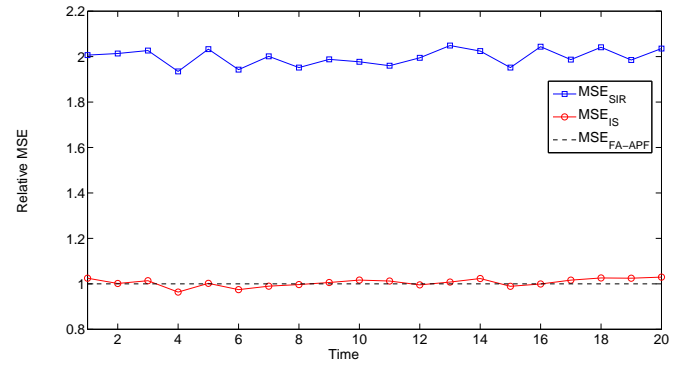
Fig. 1. Relative Mean Square Error

[3] M. K. Pitt and N. Shephard, "Filtering via simulation : Auxiliary particle filter," *J. Am. Statistical Association*, vol. 94, no. 446, pp. 550–99, june 1999.

[4] P. Fearnhead, "Computational methods for complex stochastic systems: A review of some alternatives to MCMC," *Statistics and Computing*, vol. 18, no. 2, pp. 151–71, 2008.

[5] H. R. Künsch, "State space and hidden Markov models," in *Complex Stochastic Systems*, O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg, Eds., vol. 87 of *Monographs on Statistics and Applied Probability*, pp. 109–73. Chapman and Hall / CRC, New York, 2001.

[6] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Tr. Sig. Proc.*, vol. 50, no. 3, 2002.

[7] P. del Moral, *Feynman-Kac formulae. Genealogical and interacting particle systems with applications*, Probability and its applications. Springer, New-York, 2004.

[8] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, Springer-Verlag, 2005.

[9] A. M. Johansen and A. Doucet, "A note on the auxiliary particle filter," *Statistics and Probability Letters*, vol. 78, no. 12, pp. 1498–1504, September 2008.

[10] É. Moulines R. Douc and J. Olsson, "Optimality of the auxiliary particle filter," *Probability and Mathematical Statistics*, vol. 29, no. 1, pp. 1–28, 2009.

[11] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, 2000.

[12] N. Whiteley and A. M. Johansen, "Recent developments in auxiliary particle filtering," in *Inference and Learning in Dynamic Models*, Barber, Cemgil, and Chiappa, Eds. 2010, Cambridge University Press.