A SEMI-EXACT SEQUENTIAL MONTE CARLO FILTERING ALGORITHM IN HIDDEN MARKOV CHAINS

Yohan Petetin, François Desbouvries

Telecom Institute / Telecom SudParis / CITI Department and CNRS UMR 5157
9 rue Charles Fourier, 91011 Evry, France
yohan.petetin, francois.desbouvries@telecom-sudparis.eu

ABSTRACT

Bayesian filtering is an important issue in Hidden Markov Chains (HMC) models. In many problems it is of interest to compute both the a posteriori filtering pdf at each time instant n and a moment Θ_n thereof. Sequential Monte Carlo (SMC) techniques, which include Particle filtering (PF) and Auxiliary PF (APF) algorithms, propagate a set of weighted particles which approximate that filtering pdf at time n, and then compute a Monte Carlo (MC) estimate of Θ_n. In this paper we show that in models where the so-called Fully Adapted APF (FA-APF) algorithm can be used such as semi-linear Gaussian state-space models, one can compute an estimate of the moment of interest at time n based only on the new observation y_n and on the set of particles at time $n-1$. This estimate does not suffer from the extra MC variation due to the sampling of new particles at time n, and is thus preferable to that based on that new set of particles, due to the Rao-Blackwell (RB) theorem. We finally extend our solution to models where the FA-APF cannot be used any longer.

1. INTRODUCTION

Let x_n (resp. y_n) be a sequence of hidden states (of observations), assumed to be real for simplicity. In this paper we do not distinguish random variables and their realizations. We assume that $\{(x_n, y_n)\}_{n\geq 0}$ is an HMC:

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

in which $p(x_i)$, say, is the pdf (w.r.t. Lebesgue measure) of x_i, $\mathbf{x}_{0:k} = \{x_i\}_{i=0}^{k}$, $f_{i|i-1}(x_i|x_{i-1})$ is the transition pdf of Markov Chain $\{x_n\}_{n\geq 0}$, and $g_i(y_i|x_i)$ is the likelihood of y_i given x_i. Assume that we are interested in computing, for all n, the moment

$$\Theta_n = \int_{\mathbb{R}} \phi(x_n) p(x_n|\mathbf{y}_{0:n}) \mathrm{d}x_n, \quad (2)$$

in which $p(x_n|\mathbf{y}_{0:n})$ (or simply $p_{n|n}$) is the conditional pdf of x_n given $\mathbf{y}_{0:n}$. Then one would think of computing (2) after $p_{n|n}$ is itself propagated via the well known formulas

$$p(x_n|\mathbf{y}_{0:n-1}) = \int f_{n|n-1}(x_n|\xi) p_{n-1|n-1}(\xi) \mathrm{d}\xi, \quad (3)$$

$$p(x_n|\mathbf{y}_{0:n}) = \frac{g_n(y_n|x_n) p(x_n|\mathbf{y}_{0:n-1})}{\int g_n(y_n|x_n) p(x_n|\mathbf{y}_{0:n-1}) \mathrm{d}x_n}. \quad (4)$$

However, as is well known (3)-(4) (and thus (2)) cannot be computed in practice so one need to use approximations. In particular, SMC methods consist in approximating $p_{n|n}$ by a set of weighted samples $\{w_n^i, x_n^i\}_{i=1}^N$ which are propagated over time; then the MC estimator

$$\widehat{\Theta}_n(\{w_n^i, x_n^i\}_{i=1}^N) = \sum_{i=1}^{N} w_n^i \phi(x_n^i). \quad (5)$$

of Θ_n is obtained as a byproduct.

In this paper we show that in some HMC models one can compute an estimator $\widehat{\Theta}_n^{\mathrm{se}}(\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n)$ of (2) (here superscript "se" stands for semi-exact), which depends only on the new observation y_n and on the set of particles at time $n-1$. This estimate remains an estimator of Θ_n at time n (thanks to the new data y_n), but does not use the particles $\{x_n\}_{i=1}^N$ which will be produced by the SMC algorithm. In other words, rather than propagating $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$ and next computing $\widehat{\Theta}_n(\{w_n^i, x_n^i\}_{i=1}^N)$ in (5), we first compute $\widehat{\Theta}_n^{\mathrm{se}}$ from $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$ and y_n, and next run the SMC algorithm. In this solution, $\widehat{\Theta}_n^{\mathrm{se}}$ is still a byproduct of an SMC algorithm (whence the term "semi"); but given y_n, $\widehat{\Theta}_n^{\mathrm{se}}$ is a deterministic function of $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$ (and not of samples $\{x_n^i\}_{i=1}^N$), whence the term "exact". The interest of $\widehat{\Theta}_n^{\mathrm{se}}$ is that it does not suffer from the extra MC variation induced by the production of the new samples $\{x_n\}_{i=1}^N$, and is thus preferable to $\widehat{\Theta}_n$ due to the RB theorem.

The paper is organized as follows. In §2 we show that the FA-APF algorithm can be replaced by a semi-exact solution which leads to a variance reduction, at least for classical functions such as $\phi(x_n) = x_n$ or $\phi(x_n) = x_n^2$, then we show that it is also valid for the Sampling Importance Resampling (SIR) algorithm with optimal importance distribution. When the FA-APF or the SIR algorithm with optimal importance distribution are no longer computable, we approximate our semi-exact solution by combining classical PF to get an approximation of the

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

filtering pdfs, with numerical approximations using local linearizations or Unscented Transformations (UT) to compute an estimate of Θ_n which does not rely on MC samples $\{w_n^i, x_n^i\}_{i=1}^N$. Finally in §3 we validate our semi-exact solution on popular models such as the ARCH and Stochastic Volatility (SV) models.

2. SEMI-EXACT FILTERING FOR SOME HMC: FA-APF IN SEMI-LINEAR GAUSSIAN MODELS

2.1. FA-APF-based estimator of Θ_n

In an HMC model (1), existing SMC techniques for computing $p_{n|n}$ include PF and APF filters. APF filters include the so-called FA-APF algorithm which we briefly recall. Assume that at time $n-1$ the set of weighted particles $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$ is an MC approximation of $p_{n-1|n-1}$. An approximation of $p_{n|n}$ is obtained by replacing in (3) and (4) $p_{n-1|n-1}$ by $\widehat{p}_{n-1|n-1} = \sum_{i=1}^N w_{n-1}^i \delta_{x_{n-1}^i}$. Then the following mixture pdf $\tilde{p}_{n|n}(.)$ is an approximation of $p_{n|n}$:

$$\tilde{p}_{n|n}(x_n) = \sum_{i=1}^N \frac{w_{n-1}^i p(y_n|x_{n-1}^i)}{\sum_{i=1}^N w_{n-1}^i p(y_n|x_{n-1}^i)} p(x_n|x_{n-1}^i, y_n), \tag{6}$$

where

$$p(y_n|x_{n-1}) = \int f_{n|n-1}(x_n|x_{n-1}) g_n(y_n|x_n) \mathrm{d}x_n \tag{7}$$

$$p(x_n|x_{n-1}, y_n) = \frac{f_{n|n-1}(x_n|x_{n-1}) \times g_n(y_n|x_n)}{p(y_n|x_{n-1})}. \tag{8}$$

By drawing N i.i.d. samples from $\tilde{p}_{n|n}$ in (6) we obtain the so-called FA-APF algorithm [1]:

FA-APF Algorithm.

Let $\widehat{p}_{n-1|n-1} = \sum_{i=1}^N w_{n-1}^i \delta_{x_{n-1}^i}$ be an SMC approximation of $p_{n-1|n-1}$. For all i, $1 \leq i \leq N$:

1. Sample $\tilde{x}_{n-1}^i \sim \sum_{i=1}^N \frac{w_{n-1}^i p(y_n|x_{n-1}^i)}{\sum_{i=1}^N w_{n-1}^i p(y_n|x_{n-1}^i)} \delta_{x_{n-1}^i}$;
2. Sample $x_n^i \sim p(x_n|\tilde{x}_{n-1}^i, y_n)$;
3. Set $w_n^i = \frac{1}{N}$;
4. Compute
$$\widehat{\Theta}_n = \frac{1}{N} \sum_{i=1}^N f(x_n^i). \tag{9}$$

Then $\widehat{p}_{n|n} = \sum_{i=1}^N w_n^i \delta_{x_n^i}$ approximates $p_{n|n}$, and $\widehat{\Theta}_n$ is an estimator of moment Θ_n defined in (2).

2.2. Semi-exact FA-APF-based estimator of Θ_n

The FA-APF-based estimator can be interpreted as the succession of three successive steps:

1. Approximate the filtering distribution $p_{n|n}$ by the random mixture $\tilde{p}_{n|n}$ defined in (6);
2. Draw i.i.d. samples $\{x_n^i\}_{i=1}^N$ from $\tilde{p}_{n|n}$ and set $\widehat{p}_{n|n} = \sum_{i=1}^N \frac{1}{N} \delta_{x_n^i}$;
3. Compute the empirical estimator $\widehat{\Theta}_n$ of Θ_n.

Since the estimation step (3) follows the sampling step (2), $\widehat{\Theta}_n$ depends on particles $\{x_n^i\}$ and thus suffers from the MC variations introduced by this sampling step. By construction, $\widehat{\Theta}_n$ is just a crude MC estimator $\widehat{\Theta}_n^{\mathrm{se}}$, defined as

$$\widehat{\Theta}_n^{\mathrm{se}} = \mathrm{E}_{\tilde{p}_{n|n}}(\phi(x)). \tag{10}$$

Given the previous set of particles $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$ and observation y_n,

$$\mathrm{E}(\widehat{\Theta}_n|\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n) = \widehat{\Theta}_n^{\mathrm{se}}, \tag{11}$$

$$\mathrm{var}(\widehat{\Theta}_n|\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n) = \frac{1}{N} \mathrm{var}_{\tilde{p}_{n|n}}(\phi(x)). \tag{12}$$

As we will see in the next section, in some situations it is possible to interchange steps (2) and (3) and use $\widehat{\Theta}_n^{\mathrm{se}}$ (which depends on the old particles $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$ and on y_n, but *no longer* on the new particles $\{x_n^i\}_{i=1}^N$) rather than $\widehat{\Theta}_n$. Of course, this semi-exact version of the FA-APF algorithm, which we now summarize, remains an SMC algorithm because particles $\{x_n^i\}_{i=1}^N$ are still necessary to compute the estimator at the next time step.

Semi-exact FA-APF Algorithm.

Let $\widehat{p}_{n-1|n-1} = \sum_{i=1}^N w_{n-1}^i \delta_{x_{n-1}^i}$ be an SMC approximation of $p_{n-1|n-1}$. Let $p_{i,n}(x_n) \stackrel{\text{def}}{=} p(x_n|x_{n-1}^i, y_n)$. For all i, $1 \leq i \leq N$:

1. Compute
$$\widehat{\Theta}_n^{\mathrm{se}} = \sum_{i=1}^N \frac{w_{n-1}^i p(y_n|x_{n-1}^i)}{\sum_{j=1}^N w_{n-1}^j p(y_n|x_{n-1}^j)} \mathrm{E}_{p_{i,n}}(\phi(x));$$
2. Sample $\tilde{x}_{n-1}^i \sim \sum_{i=1}^N \frac{w_{n-1}^i p(y_n|x_{n-1}^i)}{\sum_{j=1}^N w_{n-1}^j p(y_n|x_{n-1}^j)} \delta_{x_{n-1}^i}$;
3. Sample $x_n^i \sim p(x_n|\tilde{x}_{n-1}^i, y_n)$;
4. Set $w_n^i = \frac{1}{N}$.

Then $\widehat{p}_{n|n} = \sum_{i=1}^N w_n^i \delta_{x_n^i}$ approximates $p_{n|n}$, and $\widehat{\Theta}_n^{\mathrm{se}}$ is an estimator of the moment Θ_n defined in (2).

2.3. Discussion

2.3.1. Rao-Blackwell properties

Of course, $\mathrm{E}(\widehat{\Theta}_n^{\mathrm{se}}|\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n) = \mathrm{E}(\widehat{\Theta}_n|\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n) = \widehat{\Theta}_n^{\mathrm{se}}$, but we have now $\mathrm{var}(\widehat{\Theta}_n^{\mathrm{se}}|\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n) = 0$ since $\widehat{\Theta}_n^{\mathrm{se}}$ does not depend on the new particles $\{x_n^i\}_{i=1}^N$. Removing the dependency on $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$, we have the following result:

Proposition 1 *Let Θ_n be the moment defined in* (2)*, and let $\widehat{\Theta}_n$ (resp. $\widehat{\Theta}_n^{\mathrm{se}}$) be the FA-APF-based estimator (resp. semi FA-APF-based estimator) of Θ_n defined in* (9) *(resp. in* (10)*). Then*

$$\begin{aligned}
\mathrm{E}(\widehat{\Theta}_n^{\mathrm{se}}|\mathbf{y}_{0:n}) &= \mathrm{E}(\widehat{\Theta}_n|\mathbf{y}_{0:n}), & (13)\\
\mathrm{var}(\widehat{\Theta}_n^{\mathrm{se}}|\mathbf{y}_{0:n}) &\leq \mathrm{var}(\widehat{\Theta}_n|\mathbf{y}_{0:n}), & (14)\\
\mathrm{E}[(\widehat{\Theta}_n^{\mathrm{se}}-\Theta_n)^2|\mathbf{y}_{0:n}] &\leq \mathrm{E}((\widehat{\Theta}_n-\Theta_n)^2|\mathbf{y}_{0:n}). & (15)
\end{aligned}$$

Proof 1

- *The first point is obvious since the two estimators have the same conditional mean, and the weighted set $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$ has the same pdf whatever the considered estimator.*

- *For the second point, we apply the RB equality:*

$$\mathrm{var}(\widehat{\Theta}_n|\mathbf{y}_{0:n}) = \mathrm{var}(\widehat{\Theta}_n^{\mathrm{se}}|\mathbf{y}_{0:n}) + \mathrm{E}(\mathrm{var}(\widehat{\Theta}_n|\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n)|\mathbf{y}_{0:n})$$

- *Finally, the sign of the difference*

$$\mathrm{E}((\widehat{\Theta}_n^{\mathrm{se}}-\Theta_n)^2|\mathbf{y}_{0:n}) - \mathrm{E}((\widehat{\Theta}_n-\Theta_n)^2|\mathbf{y}_{0:n}) = \mathrm{var}(\widehat{\Theta}_n^{\mathrm{se}}|\mathbf{y}_{0:n}) - \mathrm{var}(\widehat{\Theta}_n|\mathbf{y}_{0:n})$$

is immediate from (14)*.*

2.3.2. Computational issues

In practice, the semi-exact algorithm requires that integral

$$\int \phi(x_n)p(x_n|x_{n-1}, y_n)\mathrm{d}x_n$$

is computable. This is impossible in general, however for some classical functions such as $\phi(x_n) = x_n$, used to give an estimator of the hidden state, the problem reduces to computing the first order moment of the sampling distribution $p(x_n|x_{n-1}, y_n)$ which happens to be available in some models (and which is also used to draw new samples $\{x_n^i\}_{i=1}^N$). Let us for instance consider the semi-linear stochastic models with additive Gaussian noise, given by

$$\begin{aligned}
x_n &= f_n(x_{n-1}) + g_n(x_{n-1})u_n & (16)\\
y_n &= h_n x_n + v_n & (17)
\end{aligned}$$

in which $\{u_n\}$ and $\{v_n\}$ are i.i.d., mutually independent and independent of x_0, $u_n \sim \mathcal{N}(0; 1)$ and $v_n \sim \mathcal{N}(0; R_n^v)$. The one-dimensional ARCH model is one such model with $f_n(x_{n-1}) = 0$, $g_n(x_{n-1}) = \sqrt{\beta_0 + \beta_1 x_{n-1}^2}$ and $h_n = 1$. In these semi-linear models, pdf $p(x_n|x_{n-1}, y_n)$ and the predictive likelihood $p(y_n|x_{n-1})$ are Gaussian and computable, so the algorithm is workable for some functions ϕ. If in particular $\phi(x) = x$, no further computation is necessary; and if $\phi(x)$ is a polynomial in x, the problem reduces to computing the first moments of the available Gaussian pdf (8). This is not the only class of functions for which the semi-exact algorithm is workable: moments according to $p(x_n|x_{n-1}, y_n)$ of functions $\phi(x_n) = \exp(ax_n)$, where a is a real constant, are also computable in such models. Note that computing $\widehat{\Theta}_n^{\mathrm{se}}$ does not require extra computational cost since in both algorithms one has to compute $p(y_n|x_{n-1})$ and the first two order moments of $p(x_n|x_{n-1}, y_n)$; in the semi-exact computation, we use them to compute directly the estimator while in the FA-APF we use them to sample new particles.

2.3.3. An alternate semi-exact solution

The variance reduction technique can indeed be applied as soon as we know how to compute $\mathrm{E}(\overline{\Theta}_n|\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n)$, where $\overline{\Theta}_n$ is an estimator of Θ_n computed via a given SMC algorithm. As we now see, the FA-APF algorithm is not the only SMC algorithm for which a semi-exact estimator is available in the ARCH model. Another popular algorithm applicable in such models is the SIR algorithm with optimal importance distribution and optional resampling step. The algorithm is a reordering of the FA-APF algorithm, and is described by the following steps:

Semi-exact optimal SIR Algorithm.

Let $\widehat{p}_{n-1|n-1} = \sum_{i=1}^N w_{n-1}^i \delta_{x_{n-1}^i}$ be an SMC approximation of $p_{n-1|n-1}$. For all i, $1 \leq i \leq N$:

1. $\tilde{x}_n^i \sim p(x_n|x_{n-1}^i, y_n)$;
2. $w_n^i \propto w_{n-1}^i p(y_n|x_{n-1}^i)$, $\sum_{i=1}^N w_n^i = 1$;
3. $\overline{\Theta}_n = \sum_{i=1}^N w_n^i \phi(\tilde{x}_n^i)$;
4. (Optional) $x_n^i \sim \sum_{i=1}^N \frac{w_{n-1}^i p(y_n|x_{n-1}^i)}{\sum_{j=1}^N w_{n-1}^j p(y_n|x_{n-1}^j)} \delta_{\tilde{x}_n^i}$.

In this algorithm, we have

$$\mathrm{E}(\overline{\Theta}_n|\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n) = \widehat{\Theta}_n^{\mathrm{se}}, \quad (18)$$

$$\mathrm{var}(\overline{\Theta}_n|\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N, y_n) = \sum_{i=1}^N \left[\frac{w_{n-1}^i p(y_n|x_{n-1}^i)}{\sum_{j=1}^N w_{n-1}^j p(y_n|x_{n-1}^j)}\right]^2 \mathrm{var}_{p_{i,n}}(\phi(x)). \quad (19)$$

Starting from a common set $\{w_{n-1}^i, x_{n-1}\}_{i=1}^N$, the estimator $\widehat{\Theta}_n^{\mathrm{se}}$ defined in Eq. (18) outperforms the Importance Sampling based estimator $\overline{\Theta}_n$ (the proof is similar to that developed in §2.3.1).

Remark 1 Until now, we have just shown that given a set $\{w_{n-1}^i, x_{n-1}^i\}$ which approximates $p_{n-1|n-1}$, it is possible in some cases to compute directly a moment according to $\tilde{p}_{n|n}$, and that this strategy is preferable to using the new set of particles $\{x_n^i\}_{i=1}^N$. However, our estimator is also based on a discrete approximation of $p_{n-1|n-1}$ and is dependent on the set $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$. We do not discuss in this paper on how to propagate the discrete approximation of the filtering distribution $p(x_n|\mathbf{y}_{0:n})$ that is to say, should we propagate it using the FA-APF algorithm or the SIR one with optimal importance distribution. This is a thorny issue since the resampling step is optional and is often done according to a particular criterion, like an estimator of the number of efficient particles [2] [3]. In addition,we know from [4] that it is not possible from an asymptotical point of view to compare the set $\{w_n^i, x_n^i\}_{i=1}^N$ produced by the SIR algorithm *before* the resampling step with that produced by the FA-APF algorithm.

However, if we consider that the resampling step is done at each time step some analysis is available. It has been discussed in [5, Ch.9] that the MC estimator of Θ_n produced by the FA-APF algorithm always outperforms (in an asymptotic normality sense) that produced by the SIR algorithm *after* the resampling step. Empirically, if we start from a set $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$, the number of different particles $\{x_n^i\}$ produced by the FA-APF is equal to N while that produced by the SIR algorithm is lower than N. We thus expect that the semi-exact estimator based on the FA-APF algorithm outperforms that built from the SIR algorithm with resampling at each time step, and this will indeed be confirmed by our simulations in §3.1.

Remark 2 In some models, $\widehat{\Theta}_n^{\mathrm{se}} = \mathrm{E}_{\tilde{p}_{n|n}}(\phi(x))$ can of course not be computed, because likelihood $p(y_n|x_{n-1})$ is not computable (and so $p(x_n|x_{n-1}, y_n)$ is not either). However, techniques such as local linearizations [6], Taylor series expansion [7], or UT [8] have been proposed for approximating $p(y_n|x_{n-1})$ and $p(x_n|x_{n-1}, y_n)$. More precisely, starting from an SMC approximation $\{w_{n-1}^i, x_{n-1}^i\}_{i=1}^N$ of the filtering pdf at time $n-1$, an approximate version of our semi-exact filtering algorithm is described by the following steps:

1. Compute an approximation of $\mathrm{E}_{\tilde{p}_{n|n}}(\phi(x))$ by using local linearizations or UT;

2. Derive an SMC approximation of $p_{n|n}$ using a classical SMC algorithm, such as an SIR or APF algorithm [1]. Previous approximations can possibly be used to derive some sampling importance distributions [7] [9]. Other SMC algorithms why optimize a given criterion are described in [10] [11].

Note that this method differs from the well known Extended Kalman Filter (EKF) or Unscented Kalman Filter (UKF) where we use numerical approximations (also based on linearizations and UT) to approximate the filtering pdf and so to deduce an estimator of Θ_n; and is also different from PF based estimators, where an estimator of Θ_n is deduced from the SMC approximation of the filtering pdf at time n. Our approximated semi-exact algorithm can thus be seen as a mixture of pure SMC and pure numerical techniques and has the advantage to avoid the propagation of errors due to numerical approximation in the EKF/UKF. In our Simulations section (see §3.3), we will discuss on the quality of the approximation of the semi-exact filtering algorithm according to the model and its parameters.

3. SIMULATIONS

We compute the empirical mean square error (MSE) at each time step, averaged on $P = 200$ simulations, and defined by $MSE(n) = \frac{1}{P}\sum_{i=1}^{P} |\widehat{\Theta}_{n|n}^j - \Theta_n^j|^2$, where $\widehat{\Theta}_{n|n}^j$ is the estimate of Θ_n^j given by one of the tested algorithm at the j-th realization and Θ_n^j is the true mean at the j-th realization given either by the Kalman Filter (KF) in the Gaussian case or a bootstrap filter with $N = 10^5$ particles otherwise.

3.1. Gaussian Model

Let us consider the following model:

$$x_{n+1} = 0.9x_n + u_n \quad (20)$$
$$y_n = x_n + v_n \quad (21)$$

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 estimate the hidden state, so $\phi(x_n) = x_n$, for all n, $1 \le n \le 50$, and compare the semi-exact estimator with the KF, which computes $\mathrm{E}_{p_{n|n}}(x)$. We run the KF, which of course here is the benchmark solution, the FA-APF algorithm, the SIR algorithm with optimal sampling distribution and resampling at each time step, the semi-exact algorithm based on the FA-APF recursion (S-FA-APF) and finally the semi-exact algorithm based on the SIR recursion with optimal sampling distribution and resampling at each time step (S-SIR). All SMC algorithms use $N = 1000$ particles. The MSE of the four estimators are computed and displayed in Fig. 1. The S-FA-APF based estimator always outperforms the FA-APF based one, and the S-SIR with optimal importance distribution based estimator, always outperforms the SIR based one. Note also that the FA-APF based estimator does not always outperform the SIR based one which is in accordance with the asymptotical analysis [4], whereas the S-FA-APF based estimator always outperforms the S-SIR based one, see Remark 1.

3.2. ARCH Model

Next, we consider the ARCH model.

$$x_{n+1} = \sqrt{\beta_0 + \beta_1 x_n^2} \times u_n \quad (22)$$
$$y_n = x_n + v_n \quad (23)$$

in which $\{u_n\}$ and $\{v_n\}$ are i.i.d, mutually independent and independent of x_0, with $x_0 \sim \mathcal{N}(0, 1)$, $u_n \sim \mathcal{N}(0, 1)$, $v_n \sim \mathcal{N}(0, R)$, $R = 3$, $\beta_0 = 1$ and $\beta_1 = 0.1$. Let us now assume that we want to estimate the hidden state x_n (so $\phi(x_n) = x_n$) and the variance of the process noise ($\phi(x_n) = \beta_0 + \beta_1 \times x_n^2$) for all n, $1 \le n \le 50$. Since $p(x_n|x_{n-1}, y_n)$ is Gaussian (see §2.3.2), it is possible to calculate both moments. We compute estimators based on the FA-APF algorithm, the S-FA-APF algorithm, with

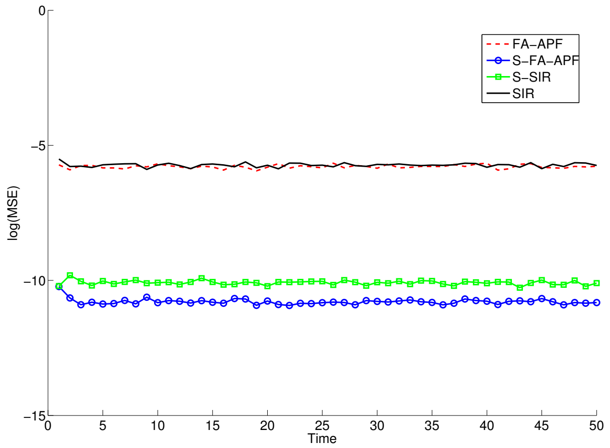

Fig. 1. MSE - Gaussian model, $R = 1$, $Q = 10$.

$N = 1000$ particles for both algorithms and the S-FA-APF algorithm with $N = 100$ particles. Results (MSE) are displayed on Fig. 2 for the estimate of x_n and Fig. 3 for the variance of the process noise.

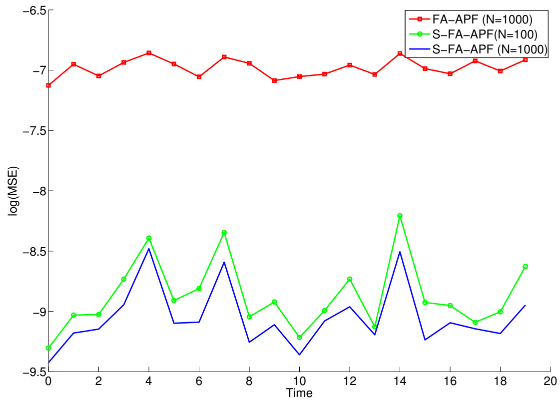

Fig. 2. MSE - ARCH Model - $\beta_0 = 1$, $\beta_1 = 0.1$ and $R = 3$ - $\phi(x_n) = x_n$

As we see, both the S-FA-APF based estimator, with $N = 1000$ particles, or $N = 100$ particles, outperform the FA-APF based one with $N = 1000$ particles: even if we start from an approximation of $p_{n-1|n-1}$ with 100 particles, it may be better to compute an estimator of Θ_n using the semi-exact FA-APF algorithm than starting from of an approximation of $p_{n-1|n-1}$ with 1000 particles and compute an estimator of Θ_n using the FA-APF algorithm. Remark however that these considerations are model and function dependent.

3.3. SV Model

Let us finally consider the following model:

$$x_{n+1} = \Phi x_n + u_n \tag{24}$$
$$y_n = \beta \exp(x_n/2) \times v_n \tag{25}$$

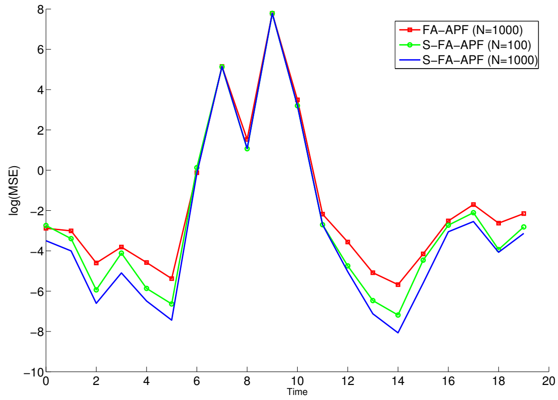

Fig. 3. MSE - ARCH Model - $\beta_0 = 1$, $\beta_1 = 0.1$ and $R = 3$ - $\phi(x_n) = \beta_0 + \beta_1 x_n^2$

in which $u_n \sim \mathcal{N}(0, \sigma^2)$ and $v_n \sim \mathcal{N}(0, 1)$. $\mathrm{E}_{\tilde{p}_{n|n}}(\phi(x))$ is not computable, whatever function ϕ, because the likelihood APF algorithms [1] [10] can be a rescue for propagating a discrete approximation of $p_{n|n}$. Such algorithms consist in sampling particles in augmented dimension:

$$(k^i, x_n^i) \sim \underbrace{\frac{w_{n-1}^j \widehat{p}(y_n|x_{n-1}^j)}{\sum_{j=1}^N w_{n-1}^j \widehat{p}(y_n|x_{n-1}^j)}}_{\Pr(k^i = j)} \widehat{p}(x_n|x_{n-1}^j, y_n),$$

then weighting them by

$$w_n^i \propto \frac{p(x_n^i|x_{n-1}^{k^i}) p(y_n|x_n^i)}{\widehat{p}(y_n|x_{n-1}^{k^i}) \widehat{p}(x_n^i|x_{n-1}^{k^i}, y_n)}, \quad \sum_{i=1}^N w_n^i = 1.$$

Although $\{x_n^i\}_{i=1}^N$ are now sampled from

$$\tilde{p}'_{n|n}(x_n) = \sum_{i=1}^N \frac{w_{n-1}^i \widehat{p}(y_n|x_{n-1}^i)}{\sum_{i=1}^N w_{n-1}^i \widehat{p}(y_n|x_{n-1}^i)} \widehat{p}(x_n|x_{n-1}^i, y_n)$$

(which differs from $\tilde{p}_{n|n}$),

$$\mathrm{E}(\sum_{i=1}^N w_n^i \phi(x_n^i)) \neq \mathrm{E}_{\tilde{p}'_{n|n}}(\phi(x))$$

because weights $\{w_n^i\}_{i=1}^N$ depend on the whole new set $\{x_n^i\}_{i=1}^N$. However, if $\tilde{p}'_{n|n}$ is a correct approximation of $\tilde{p}_{n|n}$ then it may be interesting to compute $\mathrm{E}_{\tilde{p}'_{n|n}}(\phi(x))$ and to compare it with the classical estimator based on the APF recursion [1]. The approximation of $p(y_n|x_{n-1})$ relies on a first order Taylor series expansion of function $\log(g_n(y_n|x_n))$ in Φx_{n-1}, and so one can consider that the approximation of $\log(g_n(y_n|x_n))$ is correct for values of x_n close to Φx_{n-1}. If the deduced approximation of $g_n(y_n|x_n)$ is noted $\hat{g}_n(y_n|x_n)$ then $\widehat{p}(y_n|x_{n-1}) = \int f_{n|n-1}(x_n|x_{n-1}) \times \hat{g}_n(y_n|x_n) \mathrm{d}x_n$, where $f_{n|n-1}(x_n|x_{n-1}) = \mathcal{N}(x_n; \Phi x_{n-1}; \sigma^2)$ is now computable. Let us assume that σ is small. Then $f_{n|n-1}(x_n|x_{n-1})$ is approximately non-null for values close of Φx_{n-1}, and for such values, $\hat{g}_n(y_n|x_n)$ is a good approximation of $g_n(y_n|x_n)$. An approximation of $p(x_n|x_{n-1}, y_n)$ deduced from (8) is given by a Gaussian pdf, see [1]. So one should get good approximations $\widehat{p}(y_n|x_{n-1})$ and $\widehat{p}(x_n|x_{n-1}, y_n)$ when σ^2 is small. In this simulation, we estimate the standard deviation of the observation noise at time n, so $\phi(x_n) = \beta\exp(x_n/2)$ for all n, $1 \leq n \leq 50$. We take $\Phi = 0.8$, $\beta = 0.6$, $\sigma^2 = 0.18$ and we compute the estimator based on the APF and that based on the approximated mixture $\tilde{p}'_{n|n}$ (which is computable since $\widehat{p}(x_n|x_{n-1}, y_n)$ is Gaussian and $\phi(x_n) = \beta\exp(x_n/2)$). Results are plotted in Fig. 4, and we indeed observe that the semi-exact estimator outperforms the APF based one, even if we compute an approximation of the semi-exact solution via mixture $\tilde{p}'_{n|n}$.

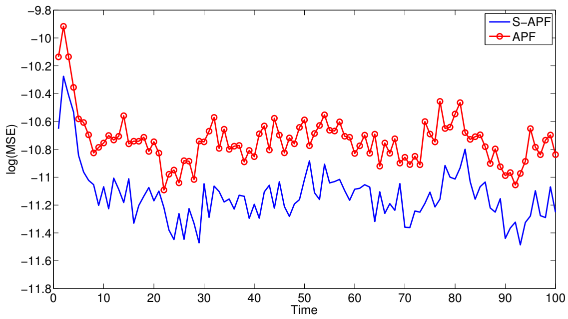

Fig. 4. MSE - Stochastic Volatility Model - $\Phi = 0.8$, $\beta = 0.6$, $Q = 0.18^2$

4. CONCLUSION

In this paper we have proposed to reduce the variance of an SMC estimator of a moment of interest of the a posteriori filtering distribution by reversing the estimation step and the sampling step. In practice, this technique can be used in models in which the FA-APF algorithm can be applied and for classical functions of interest. An approximate implementation has also been proposed when the FA-APF algorithm is not directly applicable. Simulations validated our approach.

5. REFERENCES

[1] M. K. Pitt and N. Shephard, "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–99, June 1999.

[2] A. Kong, J. S. Liu, and W. H. Wong, "Sequential Imputations and Bayesian Missing Data Problems," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 278–88, March 1994.

[3] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputation," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 567–76, June 1995.

[4] 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.

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

[6] 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.

[7] S. Saha, P. K. Manda, Y. Boers, H. Driessen, and A. Bagchi, "Gaussian Proposal density using moment matching in SMC methods." *Statistics and Computing*, vol. 19-2, pp. 203–208, 2009.

[8] S. Julier and J. Uhlmann, "Unscented Filtering and Nonlinear Estimation," in *Proceedings of the IEEE*, vol. 92, March 2004, pp. 401–422.

[9] R. van der Merwe, A. Doucet, N. de Freitas, and E. Wan, "The Unscented Particle Filter." *Advances in Neural Information Processing Systems*, 2000.

[10] J. Cornebise, É. Moulines, and J. Olsson, "Adaptive Methods for Sequential Importance Sampling with Application to State-Space Models," *Statistics and Computing*, vol. 18, no. 4, pp. 461–480, 2008.

[11] R. Douc, E. Moulines, and J. Olsson, "Optimality of the Auxiliary Particle Filter," *Probability and Mathematical Statistics*, vol. 29, no. 1, pp. 1–28, 2009.