A Note on Auxiliary Particle Filters

Adam M. Johansen [a,*], Arnaud Doucet [b]

[a] *Department of Mathematics, University of Bristol, UK*

[b] *Departments of Statistics & Computer Science, University of British Columbia, Vancouver, B.C., Canada*

Abstract

The Auxiliary Particle Filter (APF) introduced by Pitt and Shephard (1999) is a very popular alternative to Sequential Importance Sampling and Resampling (SISR) algorithms to perform inference in state-space models. We propose a novel interpretation of the APF as an SISR algorithm. This interpretation allows us to present simple guidelines to ensure good performance of the APF and the first convergence results for this algorithm. Additionally, we show that, contrary to popular belief, the asymptotic variance of APF-based estimators is not always smaller than those of the corresponding SISR estimators – even in the 'perfect adaptation' scenario.

Key words: Optimal filtering, Particle filtering, State-space models.

1 Introduction

Let $t = 1, 2, ...$ denote a discrete-time index. Consider an unobserved $\mathcal{X}-$valued Markov process $\{X_t\}_{t\geq 1}$ such that $X_1 \sim \mu(\cdot)$ and $X_t|(X_{t-1} = x_{t-1}) \sim f(\cdot|x_{t-1})$ where $f(\cdot|x_{t-1})$ is the homogeneous transition density with respect to a suitable dominating measure. The observations $\{Y_t\}_{t\geq 1}$ are conditionally independent given $\{X_t\}_{t\geq 1}$ and distributed according to $Y_t|(X_t = x_t) \sim g(\cdot|x_t)$.

For any sequence $\{z_t\}_{t\geq 1}$, we use the notation $z_{i:j} = (z_i, z_{i+1}, ..., z_j)$. In numerous applications, we are interested in estimating recursively in time the

* Corresponding author.
Email addresses: adam.johansen@bristol.ac.uk (Adam M. Johansen), arnaud@stat.ubc.ca (Arnaud Doucet).

sequence of posterior distributions $\{p(x_{1:t}|y_{1:t})\}_{t\geq 1}$ given by

$$p(x_{1:t}|y_{1:t}) \propto \mu(x_1)\, g(y_1|x_1) \prod_{k=2}^{t} f(x_k|x_{k-1})\, g(y_k|x_k). \tag{1}$$

When the model is linear Gaussian [1], the posterior distributions are Gaussian and their statistics can be computed using Kalman techniques. For non-linear non-Gaussian methods, these distributions do not typically admit a closed-form and it is necessary to employ numerical approximations. Recently, the class of Sequential Monte Carlo (SMC) methods - also known as particle filters - has emerged to solve this problem; see [6,11] for a review of the literature. Two classes of methods are primarily used: Sequential Importance Sampling and Resampling (SISR) algorithms [3,11,5] and Auxiliary Particle Filters (APF) [12,1,13].

In the literature, the APF methodology is always presented as being significantly different to the SISR methodology. It was originally introduced in [12] using auxiliary variables – hence its name. Several improvements were proposed to reduce its variance [1,13]. In [8, p. 141], the APF is presented without introducing any auxiliary variable and also reinterpreted as an SISR algorithm. However, this SISR algorithm is non-standard as it relies on a proposal distribution at time t on the path space $\mathcal{X}^t$ which is dependent on all the paths sampled previously.

We study here the version of the APF presented in [1] which only includes one resampling step at each time instance. Experimentally this version outperforms the original two-stage resampling algorithm proposed in [12] and is widely used; see [1] for a comparison of both approaches and [7] for an application to partially-observed diffusions. We propose a novel interpretation of this APF as a *standard* SISR algorithm which we believe has two principal advantages over previous derivations/interpretations. First, it allows us to give some simple guidelines to ensure good performance of the APF. These guidelines differ from many practical implementations of the APF and explain some of the poor performance reported in the literature. Second, there is no convergence result available for the APF in the literature whereas there are numerous results available for SISR algorithms; see [3] for a thorough treatment. Via this novel interpretation, we can easily adapt these results to the APF. We present here the asymptotic variance associated with APF-based estimators and show that this asymptotic variance is not necessarily lower than that of the corresponding standard SISR-based estimators – even in the 'perfect adaptation' case which is discussed further below.

[1] In the sense that μ, f and g are all Gaussian distributions with means provided by a linear function of the conditioning arguments.

2 SISR and APF

2.1 A Generic SISR algorithm

Consider an arbitrary sequence of probability distributions $\{\pi_t(x_{1:t})\}_{t\geq 1}$. To sample sequentially from these distributions, the SISR algorithm introduces at time t an importance distribution $q_t(x_t|\, x_{t-1})$ to impute X_t (and $q_1(x_1)$ at time 1). Note that it is possible to use a distribution $q_t(x_t|\, x_{1:t-1})$ but this additional freedom is not useful for the optimal filtering applications discussed here. The SISR algorithm proceeds as follows; see for example [6], [11, chapter 3] for variations:

At time 1.

Sampling Step

For $i = 1:N$, sample $X_{1,1}^{(i)} \sim q_1(\cdot)$.

Resampling Step

For $i = 1:N$, compute $w_1\left(X_{1,1}^{(i)}\right) = \dfrac{\pi_1\left(X_{1,1}^{(i)}\right)}{q_1\left(X_{1,1}^{(i)}\right)}$ and $W_1^{(i)} = \dfrac{w_1\left(X_{1,1}^{(i)}\right)}{\sum_{j=1}^N w_1\left(X_{1,1}^{(j)}\right)}$.

For $i = 1:N$, sample $\check{X}_{1,1}^{(i)} \sim \sum_{j=1}^N W_1^{(j)} \delta_{X_{1,1}^{(j)}}(dx_1)$, where $\delta_x(\cdot)$ denotes the singular distribution located at x.

At time t, $t \geq 2$.

Sampling Step

For $i = 1:N$, sample $X_{t,t}^{(i)} \sim q_t\left(\cdot|\, \check{X}_{t-1,t-1}^{(i)}\right)$.

Resampling Step

For $i = 1:N$, compute $w_t\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right) = \dfrac{\pi_t\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}{\pi_{t-1}\left(\check{X}_{1:t-1,t-1}^{(i)}\right) q_t\left(X_{t,t}^{(i)} \middle| \check{X}_{t-1,t-1}^{(i)}\right)}$

and $W_t^{(i)} = \dfrac{w_t\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}{\sum_{j=1}^N w_t\left(\check{X}_{1:t-1,t-1}^{(j)}, X_{t,t}^{(j)}\right)}$.

For $i = 1:N$, sample $\check{X}_{1:t,t}^{(i)} \sim \sum_{j=1}^N W_t^{(j)} \delta_{\left(\check{X}_{1:t-1,t-1}^{(j)}, X_{t,t}^{(j)}\right)}(dx_{1:t})$.

The empirical measure

$$\widehat{\rho}_t^N\left(dx_{1:t}\right) = \frac{1}{N}\sum_{i=1}^{N}\delta_{\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}\left(dx_{1:t}\right)$$

is an approximation of $\pi_{t-1}\left(x_{1:t-1}\right) q_t\left(x_t \middle| x_{t-1}\right)$ whereas

$$\widehat{\pi}_t^N\left(dx_{1:t}\right) = \sum_{i=1}^{N} W_t^{(i)}\delta_{\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}\left(dx_{1:t}\right)$$

is an approximation of $\pi_t\left(x_{1:t}\right)$.

Whilst, in practice, one may also wish to employ a lower variance resampling strategy such as residual resampling and to use it only when some criterion indicates that it is necessary, results of the sort presented here are sufficient to guide the design of particular algorithms and the additional complexity involved in considering more general scenarios serves largely to produce substantially more complex expressions which obscure the important points.

2.2 APF as an SISR algorithm

The standard SISR algorithm for filtering corresponds to the case in which we set $\pi_t\left(x_{1:t}\right) = p\left(x_{1:t} \middle| y_{1:t}\right)$. In this case, for any test function $\varphi_t : \mathcal{X}^t \to \mathbb{R}$, we estimate $\overline{\varphi}_t = \int \varphi_t\left(x_{1:t}\right) p\left(x_{1:t} \middle| y_{1:t}\right) dx_{1:t}$ by

$$\widehat{\varphi}_{t,SISR}^N = \int \varphi_t\left(x_{1:t}\right) \widehat{\pi}_t^N\left(dx_{1:t}\right) = \sum_{i=1}^{N} W_t^{(i)} \varphi\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right). \tag{2}$$

The APF described in [1] corresponds to the case where we select

$$\pi_t\left(x_{1:t}\right) = \widehat{p}\left(x_{1:t} \middle| y_{1:t+1}\right) \propto p\left(x_{1:t} \middle| y_{1:t}\right) \widehat{p}\left(y_{t+1} \middle| x_t\right) \tag{3}$$

with $\widehat{p}\left(y_{t+1} \middle| x_t\right)$ an approximation of

$$p\left(y_{t+1} \middle| x_t\right) = \int g\left(y_{t+1} \middle| x_{t+1}\right) f\left(x_{t+1} \middle| x_t\right) dx_{t+1}$$

if $p\left(y_{t+1} \middle| x_t\right)$ is not known analytically. As the APF does not approximate $p\left(x_{1:t} \middle| y_{1:t}\right)$ directly, we need to use importance sampling to estimate $\overline{\varphi}_t$. We use the importance distribution $\pi_{t-1}\left(x_{1:t-1}\right) q_t\left(x_t \middle| x_{t-1}\right)$ whose approximation $\widehat{\rho}_t^N\left(dx_{1:t}\right)$ is obtained after the sampling step. The resulting estimate is given by

$$\widehat{\varphi}_{t,APF}^N = \sum_{i=1}^{N} \widetilde{W}_t^{(i)} \varphi_t\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right) \tag{4}$$

where

$$\widetilde{W}_t^{(i)} = \frac{\widetilde{w}_t\left(\check{X}_{t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)}{\sum_{j=1}^{N} \widetilde{w}_t\left(\check{X}_{t-1,t-1}^{(j)}, X_{t,t}^{(j)}\right)}$$

and

$$\widetilde{w}_t\left(x_{t-1:t}\right) = \frac{p\left(x_{1:t}\middle|\, y_{1:t}\right)}{\pi_{t-1}\left(x_{1:t-1}\right) q_t\left(x_t\middle|\, x_{t-1}\right)} \propto \frac{g\left(y_t\middle|\, x_t\right) f\left(x_t\middle|\, x_{t-1}\right)}{\widehat{p}\left(y_t\middle|\, x_{t-1}\right) q_t\left(x_t\middle|\, x_{t-1}\right)}. \tag{5}$$

In both cases, we usually select $q_t\left(x_t\middle|\, x_{t-1}\right)$ as an approximation to

$$p\left(x_t\middle|\, y_t, x_{t-1}\right) = \frac{g\left(y_t\middle|\, x_t\right) f\left(x_t\middle|\, x_{t-1}\right)}{p\left(y_t\middle|\, x_{t-1}\right)}.$$

This distribution is often referred to as the optimal importance distribution [6]. When it is possible to select $q_t\left(x_t\middle|\, x_{t-1}\right) = p\left(x_t\middle|\, x_{t-1}, y_t\right)$ and $\widehat{p}\left(y_t\middle|\, x_{t-1}\right) = p\left(y_t\middle|\, x_{t-1}\right)$, we obtain the so-called 'perfect adaptation' case [12]. In this case, the APF takes a particularly simple form as the importance weights (5) are all equal. This can be interpreted as a standard SISR algorithm where the order of the sampling and resampling steps is interchanged. It is widely believed that this strategy yields estimates with a necessarily smaller variance as it increases the number of distinct particles at time t. We will show that this is not necessarily the case.

2.3 APF Settings

It is well-known in the literature that we should select for $q_t\left(x_t\middle|\, x_{t-1}\right)$ a distribution with thicker tails than $p\left(x_t\middle|\, y_t, x_{t-1}\right)$. However, this simple reinterpretation of the APF shows that we should also select a distribution $\widehat{p}\left(x_{1:t-1}\middle|\, y_{1:t}\right)$ with thicker tails than $p\left(x_{1:t-1}\middle|\, y_{1:t}\right)$ as $\widehat{p}\left(x_{1:t-1}\middle|\, y_{1:t}\right)$ is used as an importance distribution to estimate $p\left(x_{1:t-1}\middle|\, y_{1:t}\right)$. Thus $\widehat{p}\left(y_t\middle|\, x_{t-1}\right)$ should be more diffuse than $p\left(y_t\middle|\, x_{t-1}\right)$. It has been suggested in the literature to set $\widehat{p}\left(y_t\middle|\, x_{t-1}\right) = g\left(y_t\middle|\, \mu\left(x_{t-1}\right)\right)$ where $\mu\left(x_{t-1}\right)$ corresponds to the mode, mean or median of $f\left(x_t\middle|\, x_{t-1}\right)$. However, this simple approximation will often yield an importance weight function (5) which is not upper bounded on $\mathcal{X} \times \mathcal{X}$ and could lead to an estimator with a large – or even infinite – variance. An alternative, and preferable approach consists of selecting an approximation $\widehat{p}\left(y_t, x_t\middle|\, x_{t-1}\right) = \widehat{p}\left(y_t\middle|\, x_{t-1}\right) \widehat{p}\left(x_t\middle|\, y_t, x_{t-1}\right)$ of the distribution $p\left(y_t, x_t\middle|\, x_{t-1}\right) = p\left(y_t\middle|\, x_{t-1}\right) p\left(x_t\middle|\, y_t, x_{t-1}\right) = g\left(y_t\middle|\, x_t\right) f\left(x_t\middle|\, x_{t-1}\right)$ such that the ratio (5) is upper bounded on $\mathcal{X} \times \mathcal{X}$ and such that it is possible to compute $\widehat{p}\left(y_t\middle|\, x_{t-1}\right)$ pointwise and to sample from $\widehat{p}\left(x_t\middle|\, y_t, x_{t-1}\right)$.

There is a wide range of sharp convergence results available for SISR algorithms [3]. We present here a Central Limit Theorem (CLT) for the SISR and the APF estimates (2) and (4), giving the asymptotic variances of these estimates. The asymptotic variance of the CLT for the SISR estimate (2) has been established several times in the literature. We present here a new representation which we believe clarifies the influence of the ergodic properties of the optimal filter on the asymptotic variance.

Proposition. Under the regularity conditions given in [2, Theorem 1] or [3, Section 9.4, pp. 300-306], we have

$$\sqrt{N}\left(\widehat{\varphi}_{t,SISR}^{N}-\overline{\varphi}_{t}\right)\rightarrow\mathcal{N}\left(0,\sigma_{SISR}^{2}\left(\varphi_{t}\right)\right),$$
$$\sqrt{N}\left(\widehat{\varphi}_{t,APF}^{N}-\overline{\varphi}_{t}\right)\rightarrow\mathcal{N}\left(0,\sigma_{APF}^{2}\left(\varphi_{t}\right)\right)$$

where '$\rightarrow$' denotes convergence in distribution and $\mathcal{N}\left(0,\sigma^{2}\right)$ is the zero-mean normal of variance σ^{2}. Moreover, at time $t=1$ we have

$$\sigma_{SISR}^{2}\left(\varphi_{1}\right)=\sigma_{APF}^{2}\left(\varphi_{1}\right)=\int\frac{p\left(x_{1}|\,y_{1}\right)^{2}}{q_{1}\left(x_{1}\right)}\left(\varphi_{1}\left(x_{1}\right)-\overline{\varphi}_{1}\right)^{2}dx_{1}$$

whereas for $t>1$

$$\begin{aligned}\sigma_{SISR}^{2}\left(\varphi_{t}\right)=&\int\frac{p\left(x_{1}|\,y_{1:t}\right)^{2}}{q_{1}\left(x_{1}\right)}\left(\int\varphi_{t}\left(x_{1:t}\right)p\left(x_{2:t}|\,y_{2:t},x_{1}\right)dx_{2:t}-\overline{\varphi}_{t}\right)^{2}dx_{1}\\&+\sum_{k=2}^{t-1}\int\frac{p\left(x_{1:k}|\,y_{1:t}\right)^{2}}{p\left(x_{1:k-1}|\,y_{1:k-1}\right)q_{k}\left(x_{k}|\,x_{k-1}\right)}\left(\int\varphi_{t}\left(x_{1:t}\right)p\left(x_{k+1:t}|\,y_{k+1:t},x_{k}\right)dx_{k+1:t}-\overline{\varphi}_{t}\right)^{2}dx_{1:k}\\&+\int\frac{p\left(x_{1:t}|\,y_{1:t}\right)^{2}}{p\left(x_{1:t-1}|\,y_{1:t-1}\right)q_{t}\left(x_{t}|\,x_{t-1}\right)}\left(\varphi_{t}\left(x_{1:t}\right)-\overline{\varphi}_{t}\right)^{2}dx_{1:t},\end{aligned}\tag{6}$$

and

$$\begin{aligned}\sigma_{APF}^{2}(\varphi_{t})=&\int\frac{p(x_{1}|y_{1:t})^{2}}{q_{1}(x_{1})}\left(\int\varphi_{t}(x_{1:t})p(x_{2:t}|y_{2:t},x_{1})dx_{2:t}-\bar{\varphi}_{t}\right)^{2}dx_{1}\\&+\sum_{k=2}^{t-1}\int\frac{p(x_{1:k}|y_{1:t})^{2}}{\hat{p}(x_{1:k-1}|y_{1:k})q_{k}(x_{k}|x_{k-1})}\left(\int\varphi_{t}(x_{1:t})p(x_{k+1:t}|y_{k+1:t},x_{k})dx_{k+1:t}-\bar{\varphi}_{t}\right)^{2}dx_{1:k}\\&+\int\frac{p(x_{1:t}|y_{1:t})^{2}}{\hat{p}(x_{1:t-1}|y_{1:t})q_{t}(x_{t}|x_{t-1})}\left(\varphi_{t}(x_{1:t})-\bar{\varphi}_{t}\right)^{2}dx_{1:t}.\end{aligned}\tag{7}$$

Sketch of Proof. Expression (6) follows from a straightforward but tedious rewriting of the expression given in [3, Section 9.4, pp. 300-306]. We do not detail these lengthy calculations here.

The variance of the estimate $\sum_{i=1}^{N} W_t^{(i)} \varphi_t\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)$ when $\pi_t(x_{1:t})$ is given by (3) is given by an expression similar to (6) but with the terms $\widehat{p}(x_{1:k}|\, y_{1:t+1})$, $\widehat{p}(x_{1:k-1}|\, y_{1:k})$ and $\widehat{p}(x_{k+1:t-1}|\, y_{k+1:t+1}, x_k)$ replacing $p(x_{1:k}|\, y_{1:t})$, $p(x_{1:k-1}|\, y_{1:k-1})$ and $p(x_{k+1:t}|\, y_{k+1:t}, x_k)$, respectively (and with $\bar{\varphi}_t$ replaced by $\int \varphi_t(x_{1:t})\hat{p}(x_{1:t}|y_{1:t+1})dx_{1:t})$). Then by the same argument as [2, Lemma A2] the variance $\sigma_{APF}^2(\varphi_t)$ is equal to the variance of $\sum_{i=1}^{N} W_t^{(i)} \varphi_t'\left(\check{X}_{1:t-1,t-1}^{(i)}, X_{t,t}^{(i)}\right)$ where

$$\varphi_t'(x_{1:t}) = \frac{p(x_{1:t}|\, y_{1:t})}{\widehat{p}(x_{1:t}|\, y_{1:t+1})}\left[\varphi_t(x_{1:t}) - \bar{\varphi}_t\right]$$

and the expression (7) follows directly. A rigorous derivation can be found in [9] or in further detail in [10]. □

Corollary. In the perfect adaptation scenario where $\widehat{p}(y_t|\, x_{t-1}) = p(y_t|\, x_{t-1})$ and $q_t(x_t|\, x_{t-1}) = p(x_t|\, y_t, x_{t-1})$, we have

$$\begin{aligned}\sigma_{APF}^2(\varphi_t) = &\int \frac{p(x_1|y_{1:t})^2}{p(x_1|y_1)}\left(\int \varphi_t(x_{1:t})p(x_{2:t}|y_{2:t}, x_1)dx_{2:t} - \bar{\varphi}_t\right)^2 dx_1 \\ &+ \sum_{k=2}^{t-1}\int \frac{p(x_{1:k}|y_{1:t})^2}{p(x_{1:k}|y_{1:k})}\left(\int \varphi_t(x_{1:t})p(x_{k+1:t}|y_{k+1:t}, x_k)dx_{k+1:t} - \bar{\varphi}_t\right)^2 dx_{1:k} \\ &+ \int p(x_{1:t}|y_{1:t})\left(\varphi_t(x_{1:t}) - \bar{\varphi}_t\right)^2 dx_{1:t}.\end{aligned}$$

Remark. The asymptotic bias for the APF can also be established by a simple adaptation of [4, Theorem 1.1]. Both the bias and variance associated with $\varphi_t(x_{1:t}) = \varphi_t(x_t)$ can be uniformly bounded in time using [4, Proposition 4.1.]; see also [2, Theorem 5].

One may interpret these variance expressions via a local error decomposition such as that of [3, Chapters 7 & 9]. The error of the particle system estimate at time t may be decomposed as a sum of differences, specifically, the difference in the estimate due to propagating forward the particle system rather than the exact solution from each time-step to the next. Summing over all such terms gives the difference between the particle system estimate and the truth. These variance expressions illustrate that, asymptotically at least, the variance follows a similar decomposition.

Each term in the variance expressions matches an importance sampling variance. Loosely, it is the variance associated with estimating the integral of a function under the smoothing distribution $p(x_{1:k}|y_{1:t})$ using as an importance distribution the last resampling distribution propagated forward according to the proposal; the functions being integrated correspond to propagating the system forward to time t using all remaining observations and then estimating the integral of φ_t. Thus, for ergodic systems in which some forgetting property holds, the early terms in this sum will decay (at least when φ_t depends only

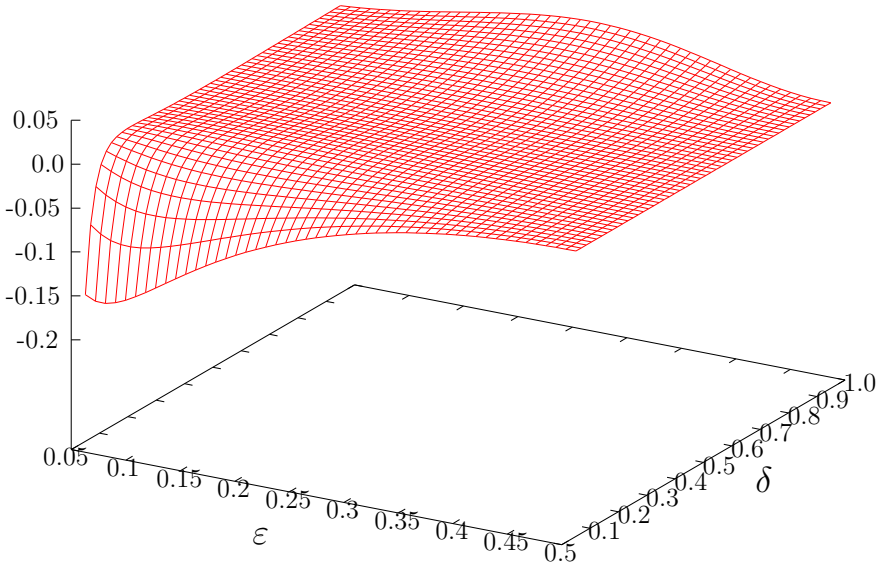

Fig. 1. Asymptotic variance difference, $\sigma^2_{APF}(\varphi_2) - \sigma^2_{SISR}(\varphi_2)$ for the example. This is negative wherever the APF outperforms SISR.

upon the final time marginal) and the system will remain well behaved over time.

3 Example

To illustrate the implications of these results, we employ the following binary state-space model with common state and observation spaces:

$$\mathcal{X} = \{0, 1\} \quad p(x_1 = 0) = 0.5 \quad p(x_t = x_{t-1}) = 1 - \delta \quad y_t \in \mathcal{X} \quad p(y_t = x_t) = 1 - \varepsilon.$$

This is an extremely simple state-space model and one could obtain the exact solution without difficulty. However, the evolution of this system from $t = 1$ to $t = 2$ provides sufficient structure to illustrate the important points and the simplicity of the model enables us to demonstrate concepts which generalise to more complex scenarios.

We consider the estimation of the function $\varphi_2(x_{1:2}) = x_2$ during the second iteration of the algorithms when the observation sequence begins $y_1 = 0, y_2 = 1$. The optimal importance distributions and the true predictive likelihood are available in this case. Additionally, the model has two parameters which are simple to interpret: δ determines how ergodic the dynamic model is (when δ

is close to 0.5 the state at time t is largely unrelated to that at time $t-1$; when it approaches 0 or 1 the two become extremely highly correlated) and ε determines how informative the observations are (when ε reaches zero, the observation at time t specifies the state deterministically, and as it approaches 0.5 it provides no information about the state).

Figure 1 shows the difference between the asymptotic variance of the APF and SISR algorithms in this setting; note that the function plotted is negative whenever the APF outperforms SISR in terms of the asymptotic variance of its estimates. A number of interesting features can be discerned. Particularly, the APF provides better estimates when δ is small, but exhibits poorer performance when $\delta \sim 1$ and $\varepsilon \sim 0.25$. When $\delta < 0.5$ the observation sequence has low probability under the prior, the APF ameliorates the situation by taking into account the predictive likelihood. The case in which ϵ and δ are both small is, unsurprisingly, that in which the APF performs best: the prior probability of the observation sequence is low, but the predictive likelihood is very concentrated.

Whilst it may appear counter-intuitive that the APF can be outperformed by SIR even in the perfect adaptation case, this can perhaps be understood by noting that perfect adaptation is simply a one-step-ahead process. The variance decomposition contains terms propagated forward from all previous times and whilst the adaptation may be beneficial at the time which it is performed it may have a negative influence on the variance at a later point. We also note that, although the APF approach does not dominate SIR, it seems likely to provide better performance in most scenarios.

Figure 2 shows experimental and asymptotic variances for the two algorithms. The displayed experimental variances were calculated as N times the empirical variance of 500 runs of each algorithm with $N = 3,000$ particles. This provides an illustration that the asymptotic results presented above do provide a useful performance guide.

4 Discussion and Extension

The main idea behind the APF, that is modifying the original sequence of target distributions to guide particles in promising regions, can be extended outside the filtering framework. Assume we are interested in a sequence of distributions $\{\pi_t(x_{1:t})\}$. Instead of using the SISR algorithm to sample from it, we use the SISR algorithm on a sequence of distributions $\{\widehat{\pi}_{t+1}(x_{1:t})\}$ where $\widehat{\pi}_{t+1}(x_{1:t})$ is an approximation of

$$\pi_{t+1}(x_{1:t}) = \int \pi_{t+1}(x_{1:t+1})\, dx_{t+1}.$$

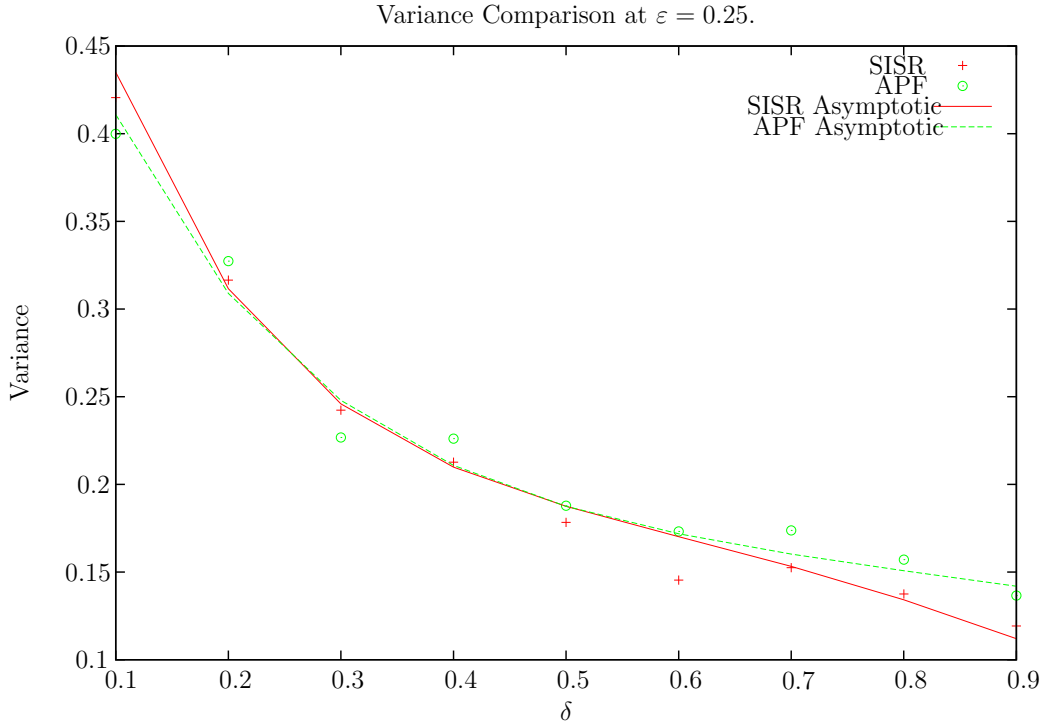

Fig. 2. Comparative variance graph: empirical and asymptotic results for the example.

We then perform inference with respect to $\pi_t(x_{1:t})$ by using importance sampling with the importance distribution $\widehat{\pi}_{t-1}(x_{1:t-1})\, q_t(x_t|\, x_{1:t-1})$ obtained after the sampling step at time t.

We also note that it has been recommended in the literature by several authors (e.g. [11, pp. 73-74]) to resample the particles not according to their normalized weights associated to $w_t^{\mathrm{SISR}}(x_{1:t}) = \frac{\pi_t(x_{1:t})}{\pi_{t-1}(x_{1:t-1}) q_t(x_t|x_{1:t-1})}$ but according to a generic score function $w_t(x_{1:t}) > 0$ at time t

$$w_t(x_{1:t}) = g\left(w_t^{\mathrm{SISR}}(x_{1:t})\right),$$

where $g : \mathbb{R}^+ \to \mathbb{R}^+$ is a monotonically increasing function; a common choice being $g(x) = x^\alpha$ where $0 < \alpha \leq 1$. To the best of our knowledge, it has never been specified clearly in the literature that this approach simply corresponds to a standard SISR algorithm for the sequence of distributions

$$\pi'_t(x_{1:t}) \propto g\left(w_t^{\mathrm{SISR}}(x_{1:t})\right) \pi_{t-1}(x_{1:t-1})\, q_t(x_t|\, x_{1:t-1}).$$

The estimates of expectations with respect to $\pi_t(x_{1:t})$ can then be computed using importance sampling. This approach is rather similar to the APF and could also be easily studied.

References

[1] Carpenter, J., Clifford, P. and Fearnhead, P. (1999) An improved particle filter for non-linear problems. *IEE proceedings - Radar, Sonar and Navigation*, **146**, 2-7.

[2] Chopin, N. (2004) Central limit theorem for sequential Monte Carlo and its application to Bayesian inference. *Ann. Statist.*, **32**, 2385-2411.

[3] Del Moral, P. (2004) *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Series Probability and Applications, New York: Springer-Verlag.

[4] Del Moral, P., Doucet, A. and Peters, G.W. (2007) Sharp propagation of chaos estimates for Feynman-Kac particle models. *Theory of Probability and Its Applications* **51**, 459-485.

[5] Doucet, A., Godsill, S.J. and Andrieu, C. (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, **10**, 197-208.

[6] Doucet, A., de Freitas, J.F.G. and Gordon, N.J. (eds.) (2001) *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

[7] Fearnhead, P., Papaspiliopoulos, O. and Roberts, G.O. (2007) Particle filters for partially-observed diffusions. Technical report, Lancaster University.

[8] Godsill, S.J. and Clapp, T. (2001) Improvement strategies for Monte Carlo particle filters, in [6], 139-158.

[9] Johansen, A.M. (2006) *Some non-standard sequential Monte Carlo methods and their applications*. PhD Thesis, Cambridge University.

[10] Johansen, A.M. and Doucet, A. (2007) Auxiliary variable sequential Monte Carlo methods. Statistics Group Technical Report, 07:09, University of Bristol. `http://www.stats.bris.ac.uk/research/stats/reports/2007/`

[11] Liu, J.S. (2001) *Monte Carlo Strategies in Scientific Computing*. New York: Springer-Verlag.

[12] Pitt, M.K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Am. Statist. Ass.*, **94**, 590-599.

[13] Pitt, M.K. and Shephard, N. (2001). Auxiliary variable based particle filters, in [6], 271–293.