# Efficient Block Sampling Strategies for Sequential Monte Carlo Methods

Arnaud DOUCET, Mark BRIERS, and Stéphane SÉNÉCAL

Sequential Monte Carlo (SMC) methods are a powerful set of simulation-based techniques for sampling sequentially from a sequence of complex probability distributions. These methods rely on a combination of importance sampling and resampling techniques. In a Markov chain Monte Carlo (MCMC) framework, block sampling strategies often perform much better than algorithms based on one-at-a-time sampling strategies if "good" proposal distributions to update blocks of variables can be designed. In an SMC framework, standard algorithms sequentially sample the variables one at a time whereas, like MCMC, the efficiency of algorithms could be improved significantly by using block sampling strategies. Unfortunately, a direct implementation of such strategies is impossible as it requires the knowledge of integrals which do not admit closed-form expressions. This article introduces a new methodology which bypasses this problem and is a natural extension of standard SMC methods. Applications to several sequential Bayesian inference problems demonstrate these methods.

**Key Words:** Block sequential Monte Carlo; Importance sampling; Markov chain Monte Carlo; Optimal filtering; Particle filtering; State-space models.

## 1. INTRODUCTION

Sequential Monte Carlo (SMC) methods are a set of flexible simulation-based methods for sampling from a sequence of probability distributions; each distribution being only known up to a normalizing constant. These methods were originally introduced in the early 1950s by physicists and have become very popular over the past few years in statistics and related fields, see Chopin (2002, 2004); Gilks and Berzuini (2001); Künsch (2005); Liu (2001); Pitt and Shephard (1999). For example, these methods are now extensively used to solve sequential Bayesian inference problems arising in econometrics, advanced signal

Arnaud Doucet is Associate Professor, Departments of Statistics and Computer Science, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada (E-mail: *arnaud@stat.ubc.ca*). Mark Briers is Research Student, Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, UK (E-mail: *mb511@cam.ac.uk*). Stéphane Sénécal is Research Associate, The Institute of Statistical Mathematics, Department of Statistical Science, Tokyo, Japan (E-mail: *steph@ism.ac.jp*).

© 2006 *American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America*

*Journal of Computational and Graphical Statistics, Volume 15, Number 3, Pages 1–19*
*DOI: 10.1198/106186006X142744*

processing or robotics; see Doucet, de Freitas, and Gordon (2001) for a comprehensive review of the literature. SMC methods approximate the sequence of probability distributions of interest using a large set of random samples, named particles. These particles are propagated over time using simple importance sampling (IS) and resampling mechanisms. Asymptotically, that is, as the number of particles goes to infinity, the convergence of these particle approximations towards the sequence of probability distributions can be ensured under very weak assumptions as discussed by Del Moral (2004). However, for practical implementations, a finite and sometimes quite restricted number of particles has to be considered. In these cases, it is crucial to design efficient sampling strategies in order to sample particles in regions of high probability mass.

A large amount of effort has been devoted to deriving improved schemes for: (1) sampling particles based on tailored importance densities (e.g., Carpenter, Clifford, and Fearnhead 1999; Doucet, Godsill, and Andrieu 2000; Guo, Wang, and Chen 2005; Liu and Chen 1998; Pitt and Shephard 1999), (2) MCMC steps to rejuvenate the particle population (e.g., Chopin 2002; Doucet, Gordon, and Krishnamurthy 2001; Fearnhead 2002; Gilks and Berzuini 2001), and (3) look-ahead techniques (e.g., Grassberger 1997; Liu 2001; Meirovitch 1985; Wang, Chen, and Guo 2002). However, tailored importance densities attempt to sample only one variable at a time, Markov chain Monte Carlo (MCMC) steps require the use of fast mixing kernels for good performance, and look-ahead techniques are computationally expensive as they typically require a "local" Monte Carlo integration for each particle. We propose here an alternative approach that allows us to extend the class of importance sampling distributions in a plethora of applications without having to perform any local Monte Carlo integration. Guidelines for the design of efficient sampling schemes based on this new framework are given. The resulting methods are natural and principled extensions of standard SMC schemes. They can be applied in a straightforward way wherever SMC methods are currently used. We demonstrate their efficiency on various optimal filtering problems.

The rest of this article is organized as follows. In Section 2, standard SMC methods are briefly reviewed and we outline their limitations. The new importance sampling approach is presented in Section 3. In Section 4, we illustrate our methodology using several optimal filtering problems. Finally, we give a brief discussion and draw conclusions in Section 5.

## 2. SEQUENTIAL MONTE CARLO METHODS

In this section we introduce the notation, briefly describe standard SMC methods, and outline their limitations—see Del Moral (2004); Doucet et al. (2000); Liu (2001) for further details. Let us consider a sequence of probability distributions $\{\pi_n\}_{n\geq 1}$ such that $\pi_n$ is defined on the product space $E_n = E^n$ and admits a density denoted $\pi_n(x_{1:n})$ with respect to a dominating measure (typically Lebesgue) where, for any general sequence $\{z_k\}$, we write $z_{i:j} = (z_i, z_{i+1}, \ldots, z_j)$. Each density is known up to a normalizing constant, that is,

$$\pi_n(x_{1:n}) = Z_n^{-1}\gamma_n(x_{1:n}),$$

where $\gamma_n : E_n \to \mathbb{R}^+$ can be evaluated pointwise whereas $Z_n$ is unknown.

SMC methods are a class of algorithms for approximately sampling sequentially from $\{\pi_n\}$; that is, first sample from $\pi_1$ then $\pi_2$ and so on. By sampling, we mean obtaining at time $n$ a collection of $N$ ($N \gg 1$) weighted random samples $\left\{W_n^{(i)}, X_{1:n}^{(i)}\right\}$, $i = 1, \ldots, N$, with $W_n^{(i)} > 0$ and $\sum_{i=1}^N W_n^{(i)} = 1$ satisfying, for any $\pi_n$-integrable function $\varphi_n$,

$$\sum_{i=1}^{N} W_n^{(i)} \varphi_n\left(X_{1:n}^{(i)}\right) \underset{N\to\infty}{\to} \int \varphi_n\left(x_{1:n}\right) \pi_n\left(x_{1:n}\right) dx_{1:n}.$$

These random samples are known as particles and are propagated through time using importance sampling and resampling mechanisms.

A popular application of SMC methods is optimal filtering, where a latent Markov process $\{X_n\}_{n\geq 1}$ is only observed through a sequence of noisy observations $\{Y_n\}_{n\geq 1}$. In this case the target distribution $\pi_n\left(x_{1:n}\right) = p\left(x_{1:n} | y_{1:n}\right)$ is the posterior distribution of $X_{1:n}$ given a realization of the observations $Y_{1:n} = y_{1:n}$; see Section 4 for additional details.

### 2.1 STANDARD SMC METHODS

We first describe the standard *sequential importance sampling resampling* (SISR) scheme. At time $n-1$, assume a set of weighted particles $\left\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\right\}$ approximating $\pi_{n-1}$ is available. Note that a random sample/particle $X_{1:n-1}^{(i)}$ represents a path from time 1 to $n-1$. The probability density of moving to $x_n$ when the current path is $x_{1:n-1}$ is denoted $q_n\left(x_n | x_{1:n-1}\right)$. The densities $\{q_n\}$ are parameters of the algorithm to be selected by the user. The algorithm proceeds as follows at time $n$.

1. Sample $X_n^{(i)} \sim q_n(\cdot | X_{1:n-1}^{(i)})$.
2. Update and normalize the weights

$$W_n^{(i)} \propto W_{n-1}^{(i)} \frac{\pi_n(X_{1:n}^{(i)})}{\pi_{n-1}(X_{1:n-1}^{(i)}) q_n(X_n^{(i)} | X_{1:n-1}^{(i)})}. \tag{2.1}$$

3. If the degeneracy of $\left\{W_n^{(i)}\right\}$ is high, resample $\left\{X_{1:n}^{(i)}\right\}$ according to $\left\{W_n^{(i)}\right\}$ to obtain $N$ unweighted particles also denoted $\left\{X_{1:n}^{(i)}\right\}$ (i.e., weights of resampled particles $W_n^{(i)} \leftarrow N^{-1}$).

The resampling step is necessary as, in most cases, the variance of the importance weights tends to increase over time. Thus, after a small number of time steps, all particles except a few have negligible weights. In the resampling operation, we associate to each particle a number of offspring proportional to its weight. Hence we focus the future computational efforts on the zones of high probability; see Doucet, de Freitas, and Gordon (2001) for several standard resampling schemes. The degeneracy of the particle representation is

typically measured using the effective sample size (ESS), as stated by Liu and Chen (1998):

$$\text{ESS} = \left( \sum_{i=1}^{N} W_n^{(i)2} \right)^{-1}. \tag{2.2}$$

The ESS takes values between 1 and $N$; if the ESS is below a given threshold, say $N/2$, then we resample; (Liu 2001, chap. 3). After the resampling step, the particles are approximately distributed according to $\pi_n$.

Expression (2.1) follows from

$$\underbrace{\frac{\pi_n(x_{1:n})}{\mu_n(x_{1:n})}}_{\text{new weight}} = \underbrace{\frac{\pi_{n-1}(x_{1:n-1})}{\mu_{n-1}(x_{1:n-1})}}_{\text{previous weight}} \underbrace{\frac{\pi_n(x_{1:n})}{\pi_{n-1}(x_{1:n-1}) q_n(x_n|x_{1:n-1})}}_{\text{incremental weight}},$$

where $\mu_n$ is the distribution of $\left\{ X_{1:n}^{(i)} \right\}$ after the sampling step; that is, if the last resampling step occurred at time $p$ ($p < n$) one has approximately

$$\mu_n \left( x_{1:n} \right) = \pi_p \left( x_{1:p} \right) \prod_{k=p+1}^{n} q_k \left( x_k | \, x_{1:k-1} \right).$$

The SISR algorithm has a computational complexity of order $O\left(N\right)$ and, for many practical applications such as optimal filtering, the calculation of the incremental weight has a fixed computational complexity. An alternative, popular SMC method is the auxiliary approach introduced by Pitt and Shephard (1999) in the context of optimal filtering.

The efficiency of the algorithms described above is highly dependent on the choice of the importance distributions $\{q_n\}$. The variance of importance sampling estimates can be shown to be approximately proportional to one plus the variance of the unnormalised importance weights; Liu (2001). In practice, the resampling step introduces correlations between particles and the variance expression is much more complex, see Chopin (2004); Del Moral (2004); Künsch (2005). However, it remains sensible to try to minimize the variance of the unnormalized importance weights appearing in the SISR algorithm.

In current approaches, the only degree of freedom we have at time $n$ is the importance distribution $q_n \left( x_n | \, x_{1:n-1} \right)$ as the paths $\left\{ X_{1:n-1}^{(i)} \right\}$ previously sampled are not modified. In this case, we are restricted to looking at the minimization of the variance of the incremental weights *conditional upon* $\left\{ X_{1:n-1}^{(i)} \right\}$. It is well known and straightforward to establish that this conditional variance is minimized for

$$q_n^{\text{opt}} \left( x_n | \, x_{1:n-1} \right) = \pi_n \left( x_n | \, x_{1:n-1} \right). \tag{2.3}$$

Using this distribution, the incremental importance weight is given by

$$\frac{\pi_n(x_{1:n})}{\pi_{n-1} \left( x_{1:n-1} \right) q_n^{\text{opt}}(x_n|x_{1:n-1})} = \frac{\pi_n \left( x_{1:n-1} \right)}{\pi_{n-1} \left( x_{1:n-1} \right)}. \tag{2.4}$$

However, it can be difficult to sample from $\pi_n \left( x_n | \, x_{1:n-1} \right)$ and/or to compute $\pi_n \left( x_{1:n-1} \right)$. Various methods have been proposed to approximate them. For example, in the optimal filtering context, $\pi_n \left( x_n | \, x_{1:n-1} \right)$ and $\pi_n \left( x_{1:n-1} \right)$ are typically approximated using standard suboptimal filtering techniques such as the extended Kalman filter; see, for example, Carpenter et al. (1999); Doucet et al. (2000); Guo et al. (2005); Pitt and Shephard (1999).

### 2.2 LIMITATIONS

Standard SMC methods suffer from several limitations. It is important to emphasize at this point that, even if the importance distribution (2.3) can be used or well approximated, this does not guarantee that the SMC algorithm will be efficient. Indeed, if the discrepancy between two consecutive distributions $\pi_n(x_{1:n-1})$ and $\pi_{n-1}(x_{1:n-1})$ is high, then the variance of the incremental weight (2.4) will be high. Consequently it will be necessary to resample very often and the particle approximation of the joint distribution

$$\widehat{\pi}_n(dx_{1:n}) = \sum_{i=1}^{N} W_n^{(i)} \delta_{X_{1:n}^{(i)}}(dx_{1:n})$$

will be unreliable. In particular, for $k << n$ the marginal distribution $\widehat{\pi}_n(dx_{1:k})$ will only be approximated by a few if not one unique particle because the algorithm will have resampled many times between times $k$ and $n$. The problem with approaches discussed until now is that only the variables $\left\{X_n^{(i)}\right\}$ are sampled at time $n$ but the paths values $\left\{X_{1:n-1}^{(i)}\right\}$ remain fixed. An obvious way to improve the algorithm would consist of not only sampling $\left\{X_n^{(i)}\right\}$ at time $n$ but also modifying the values of the paths over a fixed lag $\left\{X_{n-L+1:n-1}^{(i)}\right\}$ for $L > 1$ in light of $\pi_n$; $L$ being fixed or upper bounded to ensure that we have a sequential algorithm. The objective of this approach is not only to sample $\left\{X_n^{(i)}\right\}$ in regions of high probability mass but also to modify the path values $\left\{X_{n-L+1:n-1}^{(i)}\right\}$ to move them towards these regions. This approach is conceptually simple. Unfortunately, we will see that a direct naive implementation of it is impossible as it would require calculating an intractable integral for each particle. In the next section we present an original approach which allows us to circumvent this problem.

## 3. EFFICIENT BLOCK SAMPLING STRATEGIES FOR SMC

### 3.1 EXTENDED IMPORTANCE SAMPLING

At time $n-1$, assume a set of weighted particles $\left\{W_{n-1}^{(i)}, X_{1:n-1}^{(i)}\right\}$ approximating $\pi_{n-1}$ is available. We propose not only to extend the current paths but also to sample again a section of their paths over a fixed lag. Let $q_n\left(x'_{n-L+1:n} \middle| x_{1:n-1}\right)$ denote the probability density of moving to $x'_{n-L+1:n}$ when the current path is $x_{1:n-1}$; that is, we sample $X_{n-L+1:n}^{\prime(i)} \sim q_n(\cdot|X_{1:n-1}^{(i)})$, construct the new paths $\left\{X_{1:n-L}^{(i)}, X_{n-L+1:n}^{\prime(i)}\right\}$, and discard $\left\{X_{n-L+1:n-1}^{(i)}\right\}$. Letting $\mu_{n-1}$ denote the distribution of $\left\{X_{1:n-1}^{(i)}\right\}$ at time $n-1$, the distribution of $\left\{X_{1:n-1}^{(i)}, X_{n-L+1:n}^{\prime(i)}\right\}$ is thus given by

$$\mu_n\left(x_{1:n-1}, x'_{n-L+1:n}\right) = \mu_{n-1}(x_{1:n-1})\, q_n(x'_{n-L+1:n}|x_{1:n-1}) \tag{3.1}$$

and hence the distribution of the paths of interest $\left\{X_{1:n-L}^{(i)}, X_{n-L+1:n}^{\prime(i)}\right\}$ is

$$\mu_n\left(x_{1:n-L}, x'_{n-L+1:n}\right) = \int \mu_n\left(x_{1:n-1}, x'_{n-L+1:n}\right) dx_{n-L+1:n-1}. \tag{3.2}$$

We would like to correct for the discrepancy between $\pi_n\left(x_{1:n-L}, x'_{n-L+1:n}\right)$ and $\mu_n\left(x_{1:n-L}, x'_{n-L+1:n}\right)$ by using importance sampling. However there are two problems with this approach. First, it is usually impossible to compute $\mu_n\left(x_{1:n-L}, x'_{n-L+1:n}\right)$ pointwise up to a normalizing constant. Second, even if it were possible, there would no longer be a simple expression such as (2.1) for the weight update.

To deal with this problem, the key idea is to perform importance sampling on the enlarged space associated with $\left\{X_{1:n-1}^{(i)}, X_{n-L+1:n}^{\prime(i)}\right\}$ as their joint probability distribution (3.1) does not involve any integral. To do this it is necessary to extend the dimensionality of the target distribution $\pi_n\left(x_{1:n-L}, x'_{n-L+1:n}\right)$ to be able to perform importance sampling. We introduce an *artificial* conditional distribution $\lambda_n\left(x_{n-L+1:n-1} | x_{1:n-L}, x'_{n-L+1:n}\right)$ that allows us to define a new extended target distribution

$$\pi_n\left(x_{1:n-L}, x'_{n-L+1:n}\right) \lambda_n\left(x_{n-L+1:n-1} | x_{1:n-L}, x'_{n-L+1:n}\right).$$

As one can see, by construction, this artificial target distribution admits the required distribution $\pi_n\left(x_{1:n-L}, x'_{n-L+1:n}\right)$ as a marginal. So if we perform IS to estimate this artificial target distribution, then marginally we will obtain an estimate of $\pi_n\left(x_{1:n-L}, x'_{n-L+1:n}\right)$. It is now easy to estimate the incremental weight using the following relation

$$\begin{aligned}&\frac{\pi_n\left(x_{1:n-L}, x'_{n-L+1:n}\right) \lambda_n\left(x_{n-L+1:n-1} | x_{1:n-L}, x'_{n-L+1:n}\right)}{\mu_n\left(x_{1:n-1}, x'_{n-L+1:n}\right)} \\ &= \frac{\pi_{n-1}\left(x_{1:n-1}\right)}{\mu_{n-1}\left(x_{1:n-1}\right)} \frac{\pi_n\left(x_{1:n-L}, x'_{n-L+1:n}\right) \lambda_n\left(x_{n-L+1:n-1} | x_{1:n-L}, x'_{n-L+1:n}\right)}{\pi_{n-1}\left(x_{1:n-1}\right) q_n(x'_{n-L+1:n} | x_{1:n-1})}\end{aligned} \tag{3.3}$$

as this does not involve any integration.

Note that in this framework, $\pi_{n-1}\left(x_{1:n-1}\right) / \mu_{n-1}\left(x_{1:n-1}\right)$ does not correspond to the importance weight calculated at time $n-1$ because we also use artificial distributions before time $n$. However, if we express the target distribution at time $n$ as $\pi_n$ multiplied by *all* the artificial distributions introduced until time $n$ then the following block SISR algorithm weights the particles consistently at time $n$; see Appendix for details.

At time $n < L$

1. Sample $X_{1:n}^{\prime(i)} \sim q_n(\cdot | X_{1:n-1}^{(i)})$.
2. Update and normalize the weights

$$W_n^{(i)} \propto W_{n-1}^{(i)} \frac{\pi_n(X_{1:n}^{\prime(i)}) \lambda_n\left(X_{1:n-1}^{(i)} \middle| X_{1:n}^{\prime(i)}\right)}{\pi_{n-1}(X_{1:n-1}^{(i)}) q_n(X_{1:n}^{\prime(i)} | X_{1:n-1}^{(i)})}.$$

3. Set $\left\{X_{1:n}^{(i)}\right\} \leftarrow \left\{X_{1:n}^{\prime(i)}\right\}$.

4. If the degeneracy of $\left\{W_n^{(i)}\right\}$ is high, resample $\left\{X_{1:n}^{(i)}\right\}$ according to $\left\{W_n^{(i)}\right\}$ to obtain $N$ unweighted particles (that is, weights of resampled particles $W_n^{(i)} \leftarrow N^{-1}$).

At time $n \geq L$

5. Sample $X_{n-L+1:n}^{\prime(i)} \sim q_n(\cdot|X_{1:n-1}^{(i)})$.

6. Update and normalize the weights

$$W_n^{(i)} \propto W_{n-1}^{(i)} \underbrace{\frac{\pi_n(X_{1:n-L}^{(i)}, X_{n-L+1:n}^{\prime(i)})\lambda_n\left(X_{n-L+1:n-1}^{(i)} \middle| X_{1:n-L}^{(i)}, X_{n-L+1:n}^{\prime(i)}\right)}{\pi_{n-1}(X_{1:n-1}^{(i)})q_n(X_{n-L+1:n}^{\prime(i)}|X_{1:n-1}^{(i)})}}_{\text{incremental weight}}. \quad (3.4)$$

7. Set $\left\{X_{1:n}^{(i)}\right\} \leftarrow \left\{X_{1:n-L}^{(i)}, X_{n-L+1:n}^{\prime(i)}\right\}$.

8. If the degeneracy of $\left\{W_n^{(i)}\right\}$ is high, resample $\left\{X_{1:n}^{(i)}\right\}$ according to $\left\{W_n^{(i)}\right\}$ to obtain $N$ unweighted particles (i.e., weights of resampled particles $W_n^{(i)} \leftarrow N^{-1}$).

We adopt the convention $\pi_0(x_{1:0}) = 1$ and $\lambda_1\left(x_{1:0}|\, x_1'\right) = 1$. This algorithm is a simple and principled extension of the standard SISR procedure. An auxiliary version of this method in the spirit of Pitt and Shephard (1999) can also be obtained. General convergence results developed for SMC methods apply in a straightforward way; see Del Moral (2004). Indeed, the only difference is that instead of sampling from the initial sequence of distributions, we now sample from a sequence of extended distributions defined in the Appendix.

Clearly the performance of the algorithm is highly dependent on the artificial distributions $\{\lambda_n\}$ and the importance distributions $\{q_n\}$. In the next subsection, we provide guidelines on how to select these distributions so as to optimise the performance of the algorithm.

### 3.2 ALGORITHMIC SETTINGS

We first address the selection of the artificial distributions $\{\lambda_n\}$. To select them, we propose to minimize the variance of the incremental importance weight appearing in (3.4). We will denote this incremental weight $w_n\left(x_{1:n-1}, x_{n-L+1:n}'\right)$.

**Proposition 1.** *The conditional distribution* $\lambda_n\left(x_{n-L+1:n-1}|\, x_{1:n-L}, x_{n-L+1:n}'\right)$ *which minimizes the variance of the incremental importance weight* $w_n\left(x_{1:n-1}, x_{n-L+1:n}'\right)$ *is given by*

$$\lambda_n^{opt}\left(x_{n-L+1:n-1}|\, x_{1:n-L}, x_{n-L+1:n}'\right) = \frac{\pi_{n-1}(x_{1:n-1})q_n(x_{n-L+1:n}'|x_{1:n-1})}{\int \pi_{n-1}(x_{1:n-1})q_n(x_{n-L+1:n}'|x_{1:n-1})dx_{n-L+1:n-1}} \quad (3.5)$$

*and in this case the incremental weight takes the form*

$$w_n^{opt}\left(x_{1:n-L}, x_{n-L+1:n}'\right) = \frac{\pi_n\left(x_{1:n-L}, x_{n-L+1:n}'\right)}{\int \pi_{n-1}(x_{1:n-1})q_n(x_{n-L+1:n}'|x_{1:n-1})dx_{n-L+1:n-1}}. \quad (3.6)$$

***Proof of Proposition 1:*** The result follows from the variance decomposition formula

$$\begin{aligned}
&\text{var}\left[w_n\left(X_{1:n-1},X'_{n-L+1:n}\right)\right] \\
&\quad = E\left[\text{var}\left[w_n\left(X_{1:n-1},X'_{n-L+1:n}\right)\middle| X_{1:n-L},X'_{n-L+1:n}\right]\right] \\
&\qquad +\text{var}\left[E\left[w_n\left(X_{1:n-1},X'_{n-L+1:n}\right)\middle| X_{1:n-L},X'_{n-L+1:n}\right]\right]. \qquad (3.7)
\end{aligned}$$

The second term on the right-hand side of (3.7) is independent of $\lambda_n\left(x_{n-L+1:n-1}\middle| x_{1:n-L}\right.$, $\left.x'_{n-L+1:n}\right)$ as

$$\begin{aligned}
&E\left[w_n\left(X_{1:n-1},X'_{n-L+1:n}\right)\middle| X_{1:n-L},X'_{n-L+1:n}\right] \\
&\quad = \int \frac{\pi_n\left(x_{1:n-L},x'_{n-L+1:n}\right)\lambda_n\left(x_{n-L+1:n-1}\middle| x_{1:n-L},x'_{n-L+1:n}\right)}{\pi_{n-1}\left(x_{1:n-1}\right)q_n(x'_{n-L+1:n}|x_{1:n-1})} \\
&\qquad \times\frac{\pi_{n-1}(x_{1:n-1})q_n(x'_{n-L+1:n}|x_{1:n-1})}{\int\pi_{n-1}(x_{1:n-1})q_n(x'_{n-L+1:n}|x_{1:n-1})dx_{n-L+1:n-1}}dx_{n-L+1:n-1} \\
&\quad = w_n^{\text{opt}}\left(x_{1:n-L},x'_{n-L+1:n}\right).
\end{aligned}$$

The term $E\left[\text{var}\left[w_n\left(X_{1:n-1},X'_{n-L+1:n}\right)\middle| X_{1:n-L+1},X'_{n-L+1:n}\right]\right]$ is equal to zero if one uses the expression (3.5) for $\lambda_n$ as in this case the incremental weight becomes independent of $x_{n-L+1:n-1}$. □

This result is intuitive and simply states that the optimal artificial distribution $\lambda_n^{\text{opt}}$ is the one that takes us back to the case where we perform importance sampling on the space where the variables $x_{n-L+1:n-1}$ are integrated out. In practice, it is typically impossible however to use $\lambda_n^{\text{opt}}$ and $w_n^{\text{opt}}$, as the marginal distribution

$$\int\pi_{n-1}(x_{1:n-1})q_n(x'_{n-L+1:n}|x_{1:n-1})dx_{n-L+1:n-1} \qquad (3.8)$$

cannot be computed in closed form. There is an important exception. If $q_n(x'_{n-L+1:n}|x_{1:n-1}) = q_n(x'_{n-L+1:n}|x_{1:n-L})$, then (3.8) does not involve an integral and

$$\lambda_n^{\text{opt}}\left(x_{n-L+1:n-1}\middle| x_{1:n-L},x'_{n-L+1:n}\right) = \pi_{n-1}(x_{n-L+1:n-1}\middle| x_{1:n-L}), \qquad (3.9)$$

$$w_n^{\text{opt}}\left(x_{1:n-L},x'_{n-L+1:n}\right) = \frac{\pi_n\left(x_{1:n-L},x'_{n-L+1:n}\right)}{\pi_{n-1}(x_{1:n-L})q_n(x'_{n-L+1:n}|x_{1:n-L})}. \qquad (3.10)$$

As is the case with standard SMC previously discussed, $\pi_{n-1}(x_{1:n-L})$ is typically unknown. However, $\lambda_n$ could be selected so as to approximate (3.9). We emphasize that even if it is not equal to (3.9), this procedure still yields asymptotically consistent estimates.

Having optimized $\lambda_n$, we now consider the distributions $\{q_n\}$ that minimise the conditional variance of the incremental importance weight (3.6).

**Proposition 2.** *The importance distribution* $q_n\left(x'_{n-L+1:n}\middle| x_{1:n-1}\right)$ *which minimizes the variance of the "*$\lambda_n$*-optimized" incremental weight* $w_n^{opt}\left(x_{1:n-1},x'_{n-L+1:n}\right)$ *conditional upon* $x_{1:n-L}$ *is given by*

$$q_n^{opt}\left(x'_{n-L+1:n}\middle| x_{1:n-1}\right) = \pi_n(x'_{n-L+1:n}\middle| x_{1:n-L}) \qquad (3.11)$$

*and in this case* $w_n^{opt}\left(x_{1:n-1,}x'_{n-L+1:n}\right)$ *satisfies*

$$w_n^{opt}\left(x_{1:n-1,}x'_{n-L+1:n}\right) = \frac{\pi_n\left(x_{1:n-L}\right)}{\pi_{n-1}(x_{1:n-L})}. \tag{3.12}$$

***Proof of Proposition 2:*** The proof is straightforward as it is easy to check that the conditional variance of $w_n^{\text{opt}}$ is equal to zero for $q_n^{\text{opt}}$ given in (3.11). The expression (3.12) follows by inserting (3.11) into (3.6). □

Note that this result is a straightforward extension of the standard case where $L = 1$ as discussed in (2.3)–(2.4).

In practice, it follows from Propositions 1 and 2 that we should aim to design importance distributions $\{q_n\}$ which approximate (3.11) and then select artificial distributions $\{\lambda_n\}$ to approximate (3.9). So, if we use an approximation $\widehat{\pi}_n(x'_{n-L+1:n}|\, x_{1:n-L})$ of $\pi_n(x'_{n-L+1:n}|\, x_{1:n-L})$ for the importance distribution, we can also use an approximation $\widehat{\pi}_{n-1}(x_{n-L+1:n-1}|\, x_{1:n-L})$ of $\pi_{n-1}(x_{n-L+1:n-1}|\, x_{1:n-L})$ of the optimal artificial distribution. In this case, the block SISR algorithm proceeds as follows at time $n$ ($n \geq L$).

1. Sample $X'^{(i)}_{n-L+1:n} \sim \widehat{\pi}_n(\cdot|X^{(i)}_{1:n-L})$.
2. Update and normalize the weights
$$W_n^{(i)} \propto W_{n-1}^{(i)} \frac{\pi_n(X^{(i)}_{1:n-L}, X'^{(i)}_{n-L+1:n})\widehat{\pi}_{n-1}\left(X^{(i)}_{n-L+1:n-1}\middle| X^{(i)}_{1:n-L}\right)}{\pi_{n-1}(X^{(i)}_{1:n-1})\widehat{\pi}_n(X'^{(i)}_{n-L+1:n}|X^{(i)}_{1:n-L})}.$$
3. Set $\left\{X^{(i)}_{1:n}\right\} \leftarrow \left\{X^{(i)}_{1:n-L}, X'^{(i)}_{n-L+1:n}\right\}$.
4. If the degeneracy of $\left\{W_n^{(i)}\right\}$ is high, resample $\left\{X^{(i)}_{1:n}\right\}$ according to $\left\{W_n^{(i)}\right\}$ to obtain $N$ unweighted particles (i.e., weights of resampled particles $W_n^{(i)} \leftarrow N^{-1}$).

### 3.3 DISCUSSION

The resample-move (RM) strategy proposed by Gilks and Berzuini (2001) (see also Doucet, Gordon, and Krishnamurthy 2001; Fearnhead 2002) is a popular alternative method to limit the degeneracy of the particle population. It can also be interpreted as sampling from a sequence of artificially extended distributions. Assume we have samples $\{W^{(i)}_{n-1}, X^{(i)}_{1:n-1}\}$ approximating $\pi_{n-1}$. At time $n$, the RM algorithm first uses a standard SISR step as described in Section 2. Then the paths between time $n - L + 1$ and $n$ are "moved" according to an MCMC kernel $q_n(x'_{n-L+1:n}|x_{1:n})$ of invariant distribution $\pi_n(x_{n-L+1:n}|x_{1:n-L})$ and their weights are not modified. This MCMC step corresponds to sampling from an extended distribution

$$\pi_n\left(x_{1:n-L}, x'_{n-L+1:n}\right)\lambda_n\left(x_{n-L+1:n}|\, x_{1:n-L}, x'_{n-L+1:n}\right)$$

where the artificial measure is given by

$$\lambda_n\left(x_{n-L+1:n}|\, x_{1:n-L}, x'_{n-L+1:n}\right) = \frac{\pi_n\left(x_{1:n}\right)q_n(x'_{n-L+1:n}|x_{1:n})}{\pi_n\left(x_{1:n-L}, x'_{n-L+1:n}\right)}.$$

In practice, one introduces a resampling step between the standard SISR step and the MCMC step if the degeneracy of the importance weights is high. If this step was not introduced, then RM would be inefficient. Indeed, even if one had a very fast mixing MCMC kernel, then the weights (2.1) would not be modified. This is suboptimal. The introduction of a resampling step mitigates this problem but, contrary to the block sampling strategies described in the previous section, RM can only limit the path degeneracy over a lag of length $L$. This is demonstrated in Section 4.

In the context of static models, SMC algorithms using a RM-type strategy have been proposed by Chopin (2002) whereas algorithms based on using alternative artificial measures have been proposed by Del Moral, Doucet, and Jasra (2006). However, in Chopin (2002) and Del Moral et al. (2006), the authors use at time $n$ an MCMC kernel of invariant distribution $\pi_n$ to sample the particles, whereas the particles are sampled here using approximations of Gibbs moves.

We believe that the new approach proposed here is simpler and is a natural extension of standard techniques corresponding to the case $L = 1$. We do not claim that these block sampling SMC methods will always outperform standard SMC. It depends entirely on the ability of the user to design good approximations of the distributions $\{\pi_n(x_{n-L+1:n}|x_{1:n-L})\}$. Similarly, in a MCMC framework, block sampling strategies will only outperform one at a time strategies if the proposal distributions to sample blocks are designed carefully. A lot of effort has been devoted to the design of efficient importance distributions/proposal distributions (e.g., Durbin and Koopman 2000; Pitt and Shephard 1997) and these methods can be directly applied to our framework.

## 4. APPLICATIONS TO OPTIMAL FILTERING

### 4.1 MODEL

In this section, we detail the application of block sampling SMC methods to optimal filtering. Consider an unobserved hidden Markov process $\{X_n\}_{n\geq 1}$ defined by

$$X_1 \sim \mu,\ X_n|X_{n-1} = x_{n-1} \sim f(\cdot|x_{n-1}).$$

We only have access to noisy observations $\{Y_n\}_{n\geq 1}$. These observations are such that conditional on $\{X_n\}_{\geq 1}$ their marginal density is given by

$$Y_n|X_n = x_n \sim g(\cdot|x_n).$$

At time $n$, the optimal estimation of the collection of states $X_{1:n}$ given a realization of the observations $Y_{1:n} = y_{1:n}$ is based on the posterior density

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

The optimal distribution (3.11) and associated importance weight (3.12) are equal to

$$\pi_n\left(x_{n-L+1:n}|\, x_{1:n-L}\right) = p\left(x_{n-L+1:n}|\, y_{n-L+1:n}, x_{n-L}\right), \tag{4.1}$$

$$\pi_n\left(x_{1:n-L}\right) = p\left(x_{1:n-L}|\, y_{1:n}\right),$$

$$\frac{\pi_n\left(x_{1:n-L}\right)}{\pi_{n-1}\left(x_{1:n-L}\right)} = \frac{p\left(x_{1:n-L}|\, y_{1:n}\right)}{p\left(x_{1:n-L}|\, y_{1:n-1}\right)} \propto p\left(y_n|\, y_{n-L+1:n-1}, x_{n-L}\right). \tag{4.2}$$

We can assess the effect of the block sampling approach on the optimal importance weights in the important case where the optimal filter forgets its initial condition exponentially; see (Del Moral 2004, chap. 4) for sufficient conditions for exponential forgetting. In this importance case, under additional assumptions, it has already been established that SMC methods converge uniformly in time in $L_p$ norm in (Del Moral 2004, chap. 7) and that the variance of the SMC approximations is also bounded uniformly in time; see (Chopin 2004, theorem 5). The following simple result shows that in this case the optimal weights (4.2) also become independent of $x_{n-L}$ as $L$ increases.

**Lemma 1.** *Assume that (for finite constants $A$, $B$ and $\alpha < 1$) $g\left(y_n|\, x_n\right) < A$ for any $x_n$ and that the optimal filter forgets its initial conditions exponentially, that is, we have*

$$\int \left|p\left(x_n|\, y_{n-L+1:n}, x_{n-L}\right) - p\left(x_n|\, y_{n-L+1:n}, x'_{n-L}\right)\right| dx_n \leq B\alpha^L$$

*for any $\left(x_{n-L}, x'_{n-L}\right)$ and any $L$. In this case the optimal importance weights satisfy for any $y_n$*

$$\left|p\left(y_n|\, y_{n-L+1:n-1}, x_{n-L}\right) - p\left(y_n|\, y_{n-L+1:n-1}, x'_{n-L}\right)\right| \leq AB\alpha^L.$$

The straightforward proof is omitted. In practice, we cannot compute these weights exactly and so use approximations instead. However, this result suggests that if we can approximate the optimal importance distribution in a satisfactory way then the variance of these weights will decrease significantly with $L$, limiting drastically the number of resampling steps necessary.

Let us consider a simple Gaussian autoregressive model

$$X_n = \alpha X_{n-1} + \sigma_v V_n,$$
$$Y_n = X_n + \sigma_w W_n$$

where $V_n \overset{\text{iid}}{\sim} \mathcal{N}\left(0,1\right)$ and $W_n \overset{\text{iid}}{\sim} \mathcal{N}\left(0,1\right)$. In this case, it is easy to establish that $p\left(x_n|\, y_{n-L+1:n}, x_{n-L}\right)$ is a Gaussian distribution with covariance independent of $\left(x_n, y_{n-L+1:n}\right)$ such that

$$\left|E\left(x_n|\, y_{n-L+1:n}, x_{n-L}\right) - E\left(x_n|\, y_{n-L+1:n}, x'_{n-L}\right)\right|$$
$$= \left(\frac{\alpha}{1+\sigma_v^2/\sigma_w^2}\right)^L \left|x_{n-L} - x'_{n-L}\right|.$$

As soon as

$$\frac{|\alpha|}{1+\sigma_v^2/\sigma_w^2} < 1$$

then $p\left(x_n|\, y_{n-L+1:n}, x_{n-L}\right)$ "forgets" its initial condition exponentially quickly. This convergence is faster when the signal to noise ratio $\sigma_v^2/\sigma_w^2$ is high and the underlying Markov process $\{X_n\}$ is mixing quickly (i.e., small $\alpha$). Although we have only discussed here the linear Gaussian case (solvable through the Kalman filter), more generally the exponential forgetting property will hold when the Markov process $\{X_n\}$ mixes quickly and/or when the observations are sufficiently informative. In such situations, we expect block sampling SMC methods to outperform significantly standard methods if "good" approximations of the optimal importance distributions can be obtained.

### 4.2 SIMULATIONS

This section discusses the application of the block sampling SMC methods to two popular problems. The first problem is a target tracking problem which has been analyzed in a number of statistical publications including Fearnhead (2002) and Gilks and Berzuini (2001). The second is for stochastic volatility models appearing in Kim, Shephard, and Chib (1998), Pitt and Shephard (1997), Pitt and Shephard (1999).

#### 4.2.1 Bearing-Only Tracking

The target is modeled using a standard constant velocity model

$$X_n = \begin{pmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{pmatrix} X_{n-1} + V_n,$$

where $V_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma)$, with $T = 1$ and

$$\Sigma = 5 \begin{pmatrix} T^3/3 & T^2/2 & 0 & 0 \\ T^2/2 & T & 0 & 0 \\ 0 & 0 & T^3/3 & T^2/2 \\ 0 & 0 & T^2/2 & T \end{pmatrix}.$$

The state vector $X_n = \begin{pmatrix} X_n^1 & X_n^2 & X_n^3 & X_n^4 \end{pmatrix}^{\mathrm{T}}$ is such that $X_n^1$ (respectively $X_n^3$) corresponds to the horizontal (respectively vertical) position of the target whereas $X_n^2$ (respectively $X_n^4$) corresponds to the horizontal (respectively vertical) velocity. One only receives observations of the bearings of the target from a sensor located at the origin

$$Y_n = \tan^{-1}\left(\frac{X_n^3}{X_n^1}\right) + W_n$$

where $W_n \overset{\text{iid}}{\sim} \mathcal{N}\left(0, 10^{-4}\right)$; that is, the observations are almost noiseless. In the simulations, the initial state $X_1$ is distributed according to a Gaussian of mean corresponding to the true initial simulated point and an identity covariance. We emphasize that these parameters are representative of real-world tracking scenarios.

To build an approximation $\widehat{p}\left(x_{n-L+1:n}|\, y_{n-L+1:n}, x_{n-L}\right)$ of the optimal importance distribution (4.1), we use the extended Kalman filter (EKF) combined with the forward

Table 1. Average Number of Resampling Steps for 100 Simulations, 100 Time Instances per Simulation Using $N = 1{,}000$ Particles

| *Filter* | *Avge. # resampling steps* |
|---|---|
| Bootstrap | 46.7 |
| SMC-EKF(1) | 44.6 |
| RML(10) | 45.2 |
| RMFL(10) | 43.3 |
| SMC-EKF(2) | 34.9 |
| SMC-EKF(5) | 4.6 |
| SMC-EKF(10) | 1.3 |

filtering/backward sampling formula described by Chib (1996) and Frühwirth-Schnatter (1994). More precisely we use

$$\widehat{p}\,(x_{n-L+1:n}|\,y_{n-L+1:n}, x_{n-L}) = \widehat{p}\,(x_n|\,y_{n-L+1:n}, x_{n-L}) \prod_{k=n-L+1}^{n-1} \widehat{p}\,(x_k|\,y_{n-L+1:k}, x_{n-L}, x_{k+1})\,, \quad (4.3)$$

where

$$\widehat{p}\,(x_k|\,y_{n-L+1:k}, x_{n-L}, x_{k+1}) = \frac{f\,(x_{k+1}|\,x_k)\,\widehat{p}\,(x_k|\,y_{n-L+1:k}, x_{n-L})}{\int f\,(x_{k+1}|\,x_k)\,\widehat{p}\,(x_k|\,y_{n-L+1:k}, x_{n-L})\,dx_k}.$$

The distributions $\{\widehat{p}\,(x_k|\,y_{n-L+1:k}, x_{n-L})\}$ are Gaussian distributions whose parameters are computed using an EKF initialized using $X_{n-L} = x_{n-L}$.

We compare the following:

• The standard bootstrap filter (see, e.g., Gordon, Salmond, and Smith 1993) which uses the prior as importance distribution,

• two resample-move algorithms as described by Gilks and Berzuini (2001), where the SISR algorithm for $L = 1$ using the EKF proposal is used followed by: (1) one at a time Metropolis-Hastings (MH) moves using an approximation of the full conditionals $p\,(x_k|\,y_k, x_{k-1}, x_{k+1})$ as a proposal over a lag $L = 10$ (algorithm RML(10)); and (2) using the EKF proposal given by (4.3) for $L = 10$ (algorithm RMFL(10)). The acceptance probabilities of those moves were between 0.5/0.6 in all cases.

• the block SISR algorithms for $L = 2, 5$, and 10 which are using the EKF proposal denoted SMC-EKF($L$).

Systematic resampling is performed whenever the ESS defined in (2.2) goes below $N/2$. The results are displayed in Table 1.

The standard algorithms—namely, bootstrap, SMC-EKF(1), RML(10), and RMFL(10)—need to resample very often as the ESS drop below $N/2$. The resample-move algorithms RML(10) and RMFL(10) suffer from the same problems as standard SMC techniques (bootstrap and SMC-EKF(1)) despite their computational complexity being similar

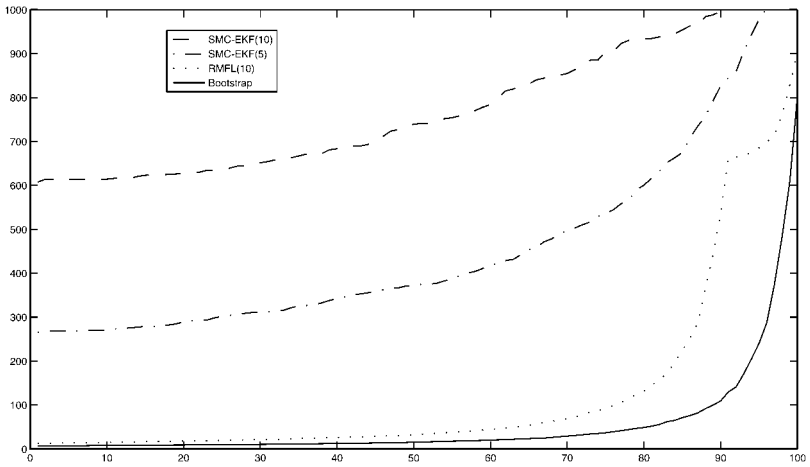


*Figure 1. Average number of unique particles* $\{X_n^{(i)}\}$ *approximating* $p(x_n|y_{1:100})$ *(y-axis) plotted against time (x-axis).*

to SMC-EKF$(10)$; this is because MCMC steps are only introduced after an EKF$(1)$ proposal has been performed. Conversely, as $L$ increases, the number of resampling steps required by SMC-EKF$(L)$ methods decreases dramatically. Consequently, the number of unique paths approximating $p\left(x_{1:100}|\, y_{1:100}\right)$ remains very large. In Figure 1, we display the average number of unique particles $\left\{X_n^{(i)}\right\}$ approximating $p\left(x_n|\, y_{1:100}\right)$. We see that using standard techniques this number rapidly decreases towards 1 as $n$ decreases whereas using the block sampling approach this decrease is much slower.

### 4.2.2 STOCHASTIC VOLATILITY

We consider the popular stochastic volatility model as described by Durbin and Koopman (2000); Kim et al. (1998); Pitt and Shephard (1997, 1999)

$$\begin{aligned} X_n &= \phi X_{n-1} + \sigma V_n,\ X_1 \sim \mathcal{N}\left(0, \frac{\sigma^2}{1-\phi^2}\right), \\ Y_n &= \beta \exp\left(X_n/2\right) W_n, \end{aligned} \tag{4.4}$$

where $V_n \overset{\text{iid}}{\sim} \mathcal{N}\left(0,1\right)$ and $W_n \overset{\text{iid}}{\sim} \mathcal{N}\left(0,1\right)$. In the SMC context, several techniques have been proposed to approximate the optimal importance distribution for $L = 1$, that is $p\left(x_n|\, y_n, x_{n-1}\right)$; Pitt and Shephard (1999). In the MCMC context (as in Pitt and Shephard 1997), methods to approximate distributions of the form $p\left(x_{n-L+1:n}|\, y_{n-L+1:n}, x_{n-L}\right)$ have been proposed but these are typically computationally intensive. We propose here a simpler alternative based on the fact that

$$\log\left(Y_n^2\right) = \log\left(\beta^2\right) + X_n + \log\left(W_n^2\right). \tag{4.5}$$

This representation has been previously used in the econometrics literature to obtain the optimal linear minimum mean square estimate of $\{X_n\}$ using the Kalman filter. We use

Table 2. Average Number of Resampling Steps for 100 Simulations using 500 Time Instances per Simulation

| *Filter* | *# particles* | *Avge. # resampling steps* |
|---|---|---|
| Bootstrap | 50000 | 176.2 |
| SMC-EKF(1) | 12000 | 127.1 |
| SMC-EKF(2) | 4000 | 80.0 |
| SMC-EKF(5) | 1600 | 11.6 |
| SMC-EKF(10) | 1000 | 0.45 |

it here to build our importance distribution. We approximate the non-Gaussian noise term $\log\left(W_n^2\right)$ with a Gaussian noise of similar mean and variance and hence obtain a linear Gaussian model approximation of (4.4)–(4.5). We then proceed in a similar fashion to the bearings-only-tracking example, by using a Gaussian approximation of the optimal distribution of the form (4.3).

The performance of our algorithms are assessed through computer simulations based on varying samples sizes to attain an approximately *equal* computational cost. We compare

- the standard bootstrap filter,
- the block SISR algorithms for $L = 1, 2, 5$, and 10 denoted SMC-EKF$(L)$.

Systematic resampling is again performed whenever the ESS goes below $N/2$. The results are displayed in Table 2 for $\sigma^2 = 0.9$, $\phi = 0.8$, and $\beta = 0.7$.

The computational complexity of the proposed approach is higher than that of standard techniques. However, as these algorithms use the observations to guide the particles in regions of high probability mass, they are much more robust to outliers than standard techniques as was clearly emphasized by Pitt and Shephard (1999). Moreover, the number of resampling steps is consequently significantly limited. This is useful if a parallel implementation is performed as the resampling operation is seen as a major bottleneck to the parallelization of SMC techniques.

Figure 2 displays the average number of unique particles $\left\{X_n^{(i)}\right\}$ approximating $p\left(x_n \middle| y_{1:500}\right)$. We see that using the standard techniques, this number decreases rather quickly as $n$ decreases. However, using the block sampling approach this decreases much more slowly. In particular, SMC-EKF(10) performs remarkably well. For $n < 400$, SMC-EKF(10) algorithm outperforms the bootstrap filter in terms of unique number of particles. It provides estimates of $p\left(x_n \middle| y_{1:500}\right)$ that are much more reliable than the bootstrap filter as $n$ decreases. Interestingly, for the same computational complexity, the bootstrap filter consistently outperforms the SMC-EKF$(1)$ algorithm for this problem. However, we emphasize here that, if outliers were present, the improvements brought by the SMC-EKF(1) algorithm and the block sampling algorithms over the bootstrap would be much higher than in these simulations.

We now apply the algorithms with $N = 1{,}000$ particles for all algorithms to the pound/dollar daily exchange rates from 1/10/81 to 28/6/85. This time series consists of

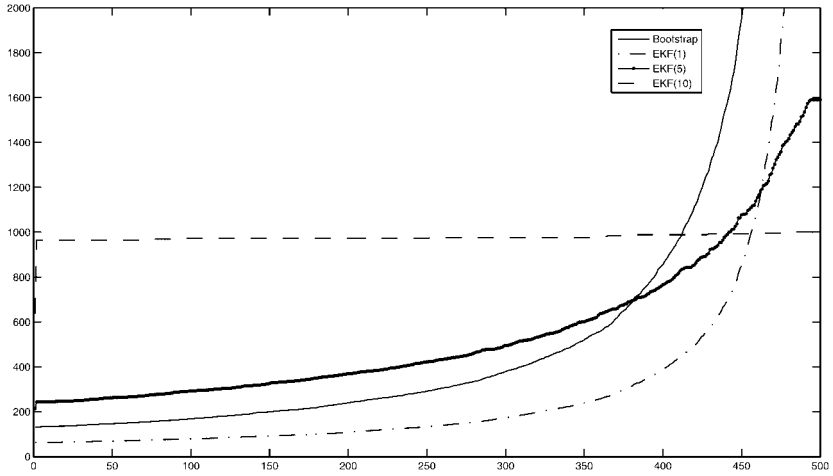


*Figure 2. Average number of unique particles $\{X_n^{(i)}\}$ approximating $p(x_n|y_{1:500})$ (y-axis) plotted against time (x-axis).*

945 data points and the parameters $\sigma = 0.1726$, $\phi = 0.9731$, $\beta = 0.6338$ are selected as in Durbin and Koopman (2000). Figure 3 displays the empirical measures approximating various marginal smoothing distributions. As expected, this approximation improves significantly as $L$ increases. Figure 4 displays SMC estimates of the posterior variances $\mathrm{var}\,[X_n|\, y_{1:945}]$. The variance estimates of the bootstrap and SMC-EKF(1) quickly decay to zero as $n$ decreases because the posterior distributions $p\,(x_n|\, y_{1:945})$ are approximated by one unique particle. The variance estimates provided by the block sampling approaches

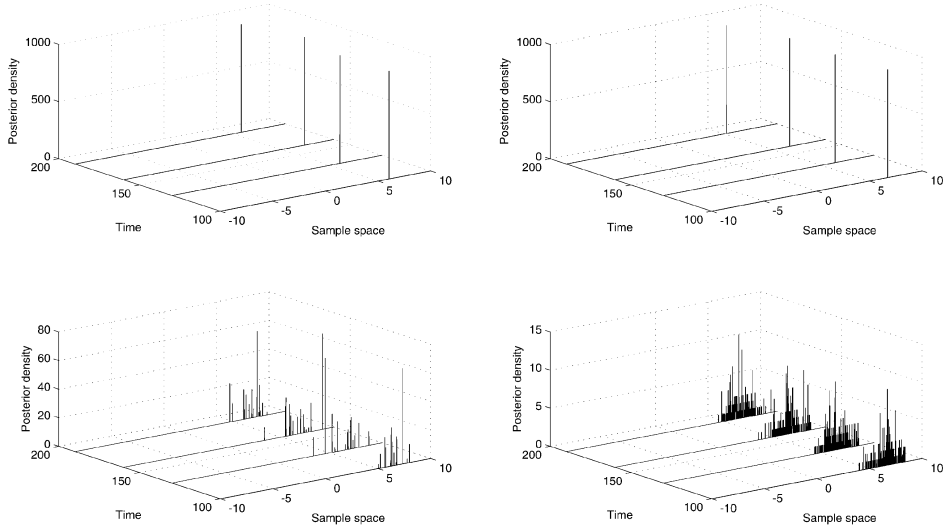


*Figure 3. Empirical measures approximating the smoothing distributions $p(x_n|y_{1:945})$ at times $n = 100$, 130, 160, 190 for bootstrap (top left), SMC-EKF(1) (top right), SMC-EKF(5) (bottom left), SMC-EKF(10) (bottom right).*

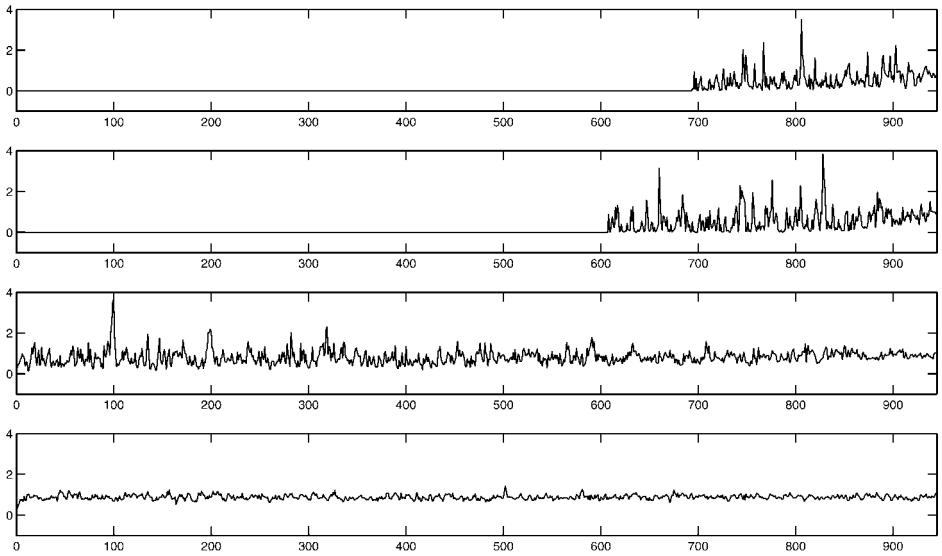


*Figure 4. SMC estimates of var $(X_n|y_{1:945})$ (y-axis) plotted against time (x-axis). Top to bottom: bootstrap, SMC-EKF(1), SMC-EKF(5), SMC-EKF(10).*

are much better. In particular SMC-EKF(10) provides variance estimates which are approximately similar as $n$ decreases; this is expected as a result of the ergodic properties of this state-space model. An MCMC run on the same dataset yields comparable estimates. This provides strong evidence that such blocking strategies can significantly limit the degeneracy of the particle population and yield much better estimates of joint distributions than standard techniques.

## 5. DISCUSSION

This article presented principled extensions of standard SMC methods that allow us to implement block sampling strategies. These methods can be applied wherever SMC methods apply. Given that the cost of block sampling schemes is higher than that of standard methods, it is difficult to assess beforehand whether it will be beneficial for a specific application. Nevertheless, the examples presented in the previous section show that it can dramatically reduce the number of resampling steps and provide a far better approximation of joint distributions than standard techniques, for a fixed computational complexity. Generally, our guidelines are that we will observe significant gains when it is possible to design a sensible approximation of the optimal importance distribution (3.11) and when the discrepancy between successive target distributions is high. In the optimal filtering framework, this situation occurs when we receive, for example, informative observations and the dynamic noise is high. This also suggests that the block sampling approach could only be used in cases where we observe a significant drop in the ESS using standard techniques.

# APPENDIX: BLOCK SAMPLING WEIGHT UPDATE DERIVATION

This appendix establishes the validity of the weights update rule (3.4). To clarify our argument, it is necessary to add a superscript to the variables; for example, $X_k^p$ corresponds to the $p$th time the random variable $X_k$ is sampled. Using such notation, a path is sampled according to

$$\begin{aligned}
X_1^1 &\sim q_1(\cdot), \\
(X_1^2, X_2^1) &\sim q_2(\cdot | X_1^1), \\
&\vdots \\
(X_1^L, X_2^{L-1}, \ldots, X_L^1) &\sim q_L(\cdot | X_1^{L-1}, \ldots, X_{L-1}^1), \\
(X_2^L, X_3^{L-1}, \ldots, X_{L+1}^1) &\sim q_{L+1}(\cdot | X_1^L, X_2^{L-1}, \ldots, X_L^1), \\
&\vdots \\
(X_{n-L+1}^L, X_{n-L+2}^{L-1} \ldots, X_n^1) &\sim q_n(\cdot | X_{1:n-L}^L, X_{n-L+1}^{L-1} \ldots, X_{n-1}^1).
\end{aligned}$$

To summarize, the importance distribution at time $n$ is of the form

$$\begin{aligned}
q_n(x_1^1, \ldots, x_n^1) = q_1(x_1^1)\, q_2(x_1^2, x_2^1 | x_1^1) \\
\times \cdots \times q_n(x_{n-L+1}^L, \ldots, x_n^1 | x_{1:n-L}^L, \ldots, x_{n-1}^1);
\end{aligned} \tag{A.1}$$

that is, at time $n$ we have sampled $L$ times the variables $x_{1:n-L+1}$ then $L-i$ times $x_{n-L+1+i}$ for $i = 1, \ldots, L$.

We now consider the following extended target distribution denoted $\widetilde{\pi}_n$

$$\begin{aligned}
&\widetilde{\pi}_n(x_1^{1:L}, \ldots, x_{n-L+1}^{1:L}, x_{n-L+2}^{1:L-1}, \ldots, x_n^1) \\
&\quad = \pi_n(x_{1:n-L+1}^L, x_{n-L+2}^{L-1}, \ldots, x_n^1)\, \lambda_2(x_1^1 | x_1^2, x_2^1) \\
&\qquad \times \cdots \times \lambda_n(x_{n-L+1}^{L-1}, \ldots, x_{n-1}^1 | x_{1:n-L}^L, x_{n-L+1}^L, \ldots, x_n^1).
\end{aligned} \tag{A.2}$$

Clearly we have

$$\underbrace{\frac{\widetilde{\pi}_n(x_1^{1:L}, \ldots, x_{n-L+1}^{1:L}, x_{n-L+2}^{1:L-1}, \ldots, x_n^1)}{q_n(x_1^{1:L}, \ldots, x_{n-L+1}^{1:L}, x_{n-L+2}^{1:L-1}, \ldots, x_n^1)}}_{\text{new weight}}$$

$$= \underbrace{\frac{\widetilde{\pi}_{n-1}(x_1^{1:L}, \ldots, x_{n-L}^{1:L}, x_{n-L+1}^{1:L-1}, \ldots, x_{n-1}^1)}{q_{n-1}(x_1^{1:L}, \ldots, x_{n-L}^{1:L}, x_{n-L+1}^{1:L-1}, \ldots, x_{n-1}^1)}}_{\text{previous weight}}$$

$$\times \underbrace{\frac{\pi_n(x_{1:n-L+1}^L, x_{n-L+2}^{L-1}, ..., x_n^1)\, \lambda_n(x_{n-L+1}^{L-1}, ..., x_{n-1}^1 | x_{1:n-L}^L, x_{n-L+1}^L, ..., x_n^1)}{\pi_{n-1}(x_{1:n-L}^L, x_{n-L+1}^{L-1}, ..., x_{n-1}^1)\, q_n(x_{n-L+1}^L, ..., x_n^1 | x_{1:n-L}^L, x_{n-L+1}^{L-1}, ..., x_{n-1}^1)}}_{\text{incremental weight}}.$$

This establishes the validity of (3.4).

*[Received April 2005. Revised January 2006.]*

## REFERENCES

Carpenter, J., Clifford, P., and Fearnhead, P. (1999), "An Improved Particle Filter for Non-linear Problems," *IEEE Proceedings—Radar, Sonar and Navigation*, 146, 2–7.

Chib, S. (1996), "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 539–552.

Chopin, N. (2002), "A Sequential Particle Filter Method for Static Models," *Biometrika*, 89, 539–552.

——— (2004), "Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference," *The Annals of Statistics*, 32, 2385–2411.

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

Del Moral, P., Doucet, A., and Jasra, A. (2006), "Sequential Monte Carlo Samplers," *Journal of the Royal Statistical Society*, Series B, 68, 411–436.

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

Doucet, A., Godsill, S., and Andrieu, C. (2000), "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, 10, 197–208.

Doucet, A., Gordon, N., and Krishnamurthy, V. (2001), "Particle Filters for State Estimation of Jump Markov Linear Systems," *IEEE Transactions on Signal Processing*, 49, 613–624.

Durbin, J., and Koopman, S. (2000), "Time Series Analysis of Non-Gaussian Observations Based on State Space Models from Both Classical and Bayesian Perspectives" (with discussion), *Journal of the Royal Statistical Society*, Series B, 62, 3–56.

Fearnhead, P. (2002), "MCMC, Sufficient Statistics and Particle Filter," *Journal of Computational and Graphical Statistics*, 11, 848–862.

Frühwirth-Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 15, 183–202.

Gilks, W., and Berzuini, C. (2001), "Following a Moving Target—Monte Carlo Inference for Dynamic Bayesian Models," *Journal of the Royal Statistical Society*, Series B, 63, 127–146.

Gordon, N., Salmond, D., and Smith, A. (1993), "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEEE Proceedings—F*, 140, 107–113.

Grassberger, P. (1997), "Pruned-Enriched Rosenbluth Method: Simulations of Theta Polymers of Chain Length up to 1,000,000," *Physical Review E*, 56, 3682–3693.

Guo, D., Wang, X., and Chen, R. (2005), "New Sequential Monte Carlo Methods for Nonlinear Dynamic Systems," *Statistics and Computing*, 15, 135–147.

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–394.

Künsch, H. (2005), "Recursive Monte Carlo filters: Algorithms and Theoretical Analysis," *The Annals of Statistics*, 33, 1983–2021.

Liu, J. (2001), *Monte Carlo Strategies in Scientific Computing*, Berlin: Springer.

Liu, J., and Chen, R. (1998), "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93, 1032–1044.

Meirovitch, H. (1985), "Scanning Method as an Unbiased Simulation Technique and its Application to the Study of Self-Attracting Random Walks," *Physical Review A*, 3699–3708.

Pitt, M., and Shephard, N. (1997), "Likelihood Analysis of non-Gaussian Measurement Time Series," *Biometrika*, 84, 653–670.

——— (1999), "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94, 590–599.

Wang, X., Chen, R., and Guo, D. (2002), "Delayed-Pilot Sampling for Mixture Kalman Filter with Applications in Fading Channels," *IEEE Transactions on Signal Processing*, 50, 241–253.