On Sequential Monte Carlo Sampling Methods for Bayesian Filtering

Arnaud Doucet (corresponding author) - Simon Godsill - Christophe Andrieu

Signal Processing Group, Department of Engineering

University of Cambridge

Trumpington Street, CB2 1PZ Cambridge, UK

Email: ad2@eng.cam.ac.uk

ABSTRACT

In this article, we present an overview of methods for sequential simulation from posterior distributions. These methods are of particular interest in Bayesian filtering for discrete time dynamic models that are typically nonlinear and non-Gaussian. A general importance sampling framework is developed that unifies many of the methods which have been proposed over the last few decades in several different scientific disciplines. Novel extensions to the existing methods are also proposed. We show in particular how to incorporate local linearisation methods similar to those which have previously been employed in the deterministic filtering literature; these lead to very effective importance distributions. Furthermore we describe a method which uses Rao-Blackwellisation in order to take advantage of the analytic structure present in some important classes of state-space models. In a final section we develop algorithms for prediction, smoothing and evaluation of the likelihood in dynamic models.

Keywords: Bayesian filtering, nonlinear non-Gaussian state space models, sequential Monte Carlo methods, importance sampling, Rao-Blackwellised estimates

I. Introduction

Many problems in applied statistics, statistical signal processing, time series analysis and econometrics can be stated in a state space form as follows. A transition equation describes the prior distribution of a hidden Markov process $\{\mathbf{x}_k; k \in \mathbb{N}\}$, the so-called hidden state process, and an observation equation describes the likelihood of the observations $\{\mathbf{y}_k; k \in \mathbb{N}\}$, k being a discrete time index. Within a Bayesian framework, all relevant information about $\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k\}$ given observations up to and including time k can be obtained from the posterior distribution $p(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k | \mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_k)$. In many applications we are interested in *estimating recursively in time* this distribution and particularly one of its marginals, the so-called filtering distribution $p(\mathbf{x}_k | \mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_k)$. Given the filtering distribution one can then routinely proceed to filtered point estimates such as the posterior mode or mean of the state. This problem is known as the Bayesian filtering problem or the optimal filtering problem. Practical applications include target tracking (Gordon et al., 1993), blind deconvolution of digital communications channels (Clapp et al., 1999)(Liu et al., 1995), estimation of stochastic volatility (Pitt et al., 1999) and digital enhancement of speech and audio signals (Godsill et al., 1998).

Except in a few special cases, including linear Gaussian state space models (Kalman filter) and hidden finite-state space Markov chains, it is impossible to evaluate these distributions analytically. From the mid 1960's, a great deal of attention has been devoted to approximating these filtering distributions, see for example (Jazwinski, 1970). The most popular algorithms, the extended Kalman filter and the Gaussian sum filter, rely on analytical approximations (Anderson et al., 1979). Interesting work in the automatic control field was carried out during the 1960's and 70's using sequential Monte Carlo (MC) integration methods, see (Akashi et al., 1975)(Handschin et. al, 1969)(Handschin 1970)(Zaritskii et al., 1975). Possibly owing to the severe computational limitations of the time these Monte Carlo algorithms have been largely neglected until recently. In the late 80's, massive increases in computational power allowed the rebirth of numerical integration methods for Bayesian filtering (Kitagawa 1987). Current research has now focused on MC integration methods, which have the great advantage of not being subject to the assumption of linearity or Gaussianity in the model, and relevant work includes (Müller 1992)(West, 1993)(Gordon et al., 1993)(Kong et al., 1994)(Liu et al., 1998).

The main objective of this article is to include in a unified framework many old and more recent algorithms proposed independently in a number of applied science areas. Both (Liu et al., 1998) and (Doucet, 1997) (Doucet, 1998) underline the central rôle of sequential importance sampling in Bayesian filtering. However, contrary to (Liu et al., 1998) which emphasizes the use of hybrid schemes combining elements of importance sampling with Markov Chain Monte Carlo (MCMC), we focus here on computationally cheaper alternatives. We describe also how it is possible to improve current existing methods via Rao-Blackwellisation for a useful class of dynamic models. Finally, we show how to extend these methods to compute the prediction and fixed-interval smoothing distributions as well as the likelihood.

The paper is organised as follows. In section 2, we briefly review the Bayesian filtering problem and classical Bayesian importance sampling is proposed for its solution. We then present a sequential version of this method which allows us to obtain a general recursive MC filter: the sequential importance sampling (SIS) filter. Under a criterion of minimum conditional variance of the importance weights, we obtain the optimal importance function for this method. Unfortunately, for numerous models of applied interest the optimal importance function leads to non-analytic importance weights, and hence we propose several suboptimal distributions and show how to obtain as special cases many of the algorithms presented in the literature. Firstly we consider local linearisation methods of either the state space model or the optimal importance function, giving some important examples. These linearisation methods seem to be a very promising way to proceed in problems of this type. Secondly we consider some simple importance functions which lead to algorithms currently known in the literature. In Section 3, a resampling scheme is used to limit practically the degeneracy of the algorithm. In Section 4, we apply the Rao-Blackwellisation method to SIS and obtain efficient hybrid analytical/MC filters. In Section 5, we show how to use the MC filter to compute the prediction and fixed-interval smoothing distributions as well as the likelihood. Finally, simulations are presented in Section 6.

II. Filtering Via Sequential Importance Sampling

A. Preliminaries: Filtering for the State Space Model

The state sequence $\{\mathbf{x}_k; k \in \mathbb{N}\}$, $\mathbf{x}_k \in \mathbb{R}^{n_\mathbf{x}}$, is assumed to be an unobserved (hidden) Markov process with initial distribution $p(\mathbf{x}_0)$ (which we subsequently denote as $p(\mathbf{x}_0|\mathbf{x}_{-1})$ for notational convenience) and transition distribution $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, where n_x is the dimension of the state vector. The observations $\{\mathbf{y}_k; k \in \mathbb{N}\}$, $\mathbf{y}_k \in \mathbb{R}^{n_\mathbf{y}}$, are conditionally independent given the process $\{\mathbf{x}_k; k \in \mathbb{N}\}$ with distribution $p(\mathbf{y}_k|\mathbf{x}_k)$ and n_y is the dimension of the observation vector. To sum up, the model is a hidden Markov (or state space) model (HMM) described by

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) \text{ for } k \geq 0 \tag{1}$$

$$p(\mathbf{y}_k|\mathbf{x}_k) \text{ for } k \geq 0 \tag{2}$$

We denote by $\mathbf{x}_{0:n} \triangleq \{\mathbf{x}_0, ..., \mathbf{x}_n\}$ and $\mathbf{y}_{0:n} \triangleq \{\mathbf{y}_0, ..., \mathbf{y}_n\}$, respectively, the state sequence and the observations up to time n. Our aim is to estimate recursively in time the distribution $p(\mathbf{x}_{0:n}|\mathbf{y}_{0:n})$ and its associated features including $p(\mathbf{x}_n|\mathbf{y}_{0:n})$ and expectations of the form

$$I(f_n) = \int f_n(\mathbf{x}_{0:n}) p(\mathbf{x}_{0:n}|\mathbf{y}_{0:n}) d\mathbf{x}_{0:n} \tag{3}$$

for any $p(\mathbf{x}_{0:n}|\,\mathbf{y}_{0:n})$-integrable $f_n : \mathbb{R}^{(n+1)\times n_{\mathbf{x}}} \to \mathbb{R}$. A recursive formula for $p(\mathbf{x}_{0:n}|\,\mathbf{y}_{0:n})$ is given by:

$$p(\mathbf{x}_{0:n+1}|\,\mathbf{y}_{0:n+1}) = p(\mathbf{x}_{0:n}|\,\mathbf{y}_{0:n}) \frac{p(\mathbf{y}_{n+1}|\,\mathbf{x}_{n+1})\, p(\mathbf{x}_{n+1}|\,\mathbf{x}_n)}{p(\mathbf{y}_{n+1}|\,\mathbf{y}_{0:n})} \tag{4}$$

The denominator of this expression cannot typically be computed analytically, thus rendering an analytic approach infeasible except in the special cases mentioned above. It will later be assumed that samples can easily be drawn from $p(\mathbf{x}_k|\,\mathbf{x}_{k-1})$ and that we can evaluate $p(\mathbf{x}_k|\,\mathbf{x}_{k-1})$ and $p(\mathbf{y}_k|\,\mathbf{x}_k)$ pointwise.

B. Bayesian Sequential Importance Sampling (SIS)

Since it is generally impossible to sample from the state posterior $p(\mathbf{x}_{0:n}|\,\mathbf{y}_{0:n})$ directly, we adopt an importance sampling (IS) approach. Suppose that samples $\left\{\mathbf{x}_{0:n}^{(i)}; i = 1, ..., N\right\}$ are drawn independently from a normalised importance function $\pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{0:n})$ which has the same support as the state posterior. Then an estimate $\widehat{I_N}(f_n)$ of the posterior expectation $I(f_n)$ is obtained using Bayesian IS (Geweke, 1989):

$$\widehat{I_N}(f_n) = \sum_{i=1}^{N} f_n\left(\mathbf{x}_{0:n}^{(i)}\right) \widetilde{w}_n^{(i)}, \qquad \widetilde{w}_n^{(i)} = \frac{w_n^{*(i)}}{\sum_{j=1}^{N} w_n^{*(j)}} \tag{5}$$

where $w_n^{*(i)} = p(\mathbf{y}_{0:n}|\,\mathbf{x}_{0:n})\, p(\mathbf{x}_{0:n}) / \pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{0:n})$ is the unnormalised importance weight. Under weak assumptions $\widehat{I_N}(f_n)$ converges to $I(f_n)$, see for example (Geweke, 1989). However, this method is not recursive. We now show how to obtain a sequential MC filter using Bayesian IS.

Suppose one chooses an importance function of the form

$$\pi(\mathbf{x}_{0:n}|\,\mathbf{y}_{0:n}) = \pi(\mathbf{x}_0|\,\mathbf{y}_0) \prod_{k=1}^{n} \pi(\mathbf{x}_k|\,\mathbf{x}_{0:k-1}, \mathbf{y}_{0:k}) \tag{6}$$

Such an importance function allows recursive evaluation in time of the importance weights as successive observations $\mathbf{y}_k$ become available. We obtain directly the sequential importance sampling filter.

Sequential Importance Sampling (SIS)

For times $k = 0, 1, 2, ...$

- For $i = 1, ..., N$, sample $\mathbf{x}_k^{(i)} \sim \pi\left(\mathbf{x}_k \middle| \mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}_{0:k}\right)$ and $\mathbf{x}_{0:k}^{(i)} \triangleq \left(\mathbf{x}_{0:k-1}^{(i)}, \mathbf{x}_k^{(i)}\right)$.

- For $i = 1, ..., N$, evaluate the importance weights up to a normalising constant:

$$w_k^{*(i)} = w_{k-1}^{*(i)} \frac{p\left(\mathbf{y}_k \middle| \mathbf{x}_k^{(i)}\right) p\left(\mathbf{x}_k^{(i)} \middle| \mathbf{x}_{k-1}^{(i)}\right)}{\pi\left(\mathbf{x}_k^{(i)} \middle| \mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}_{0:k}\right)} \tag{7}$$

- For $i = 1, ..., N$, normalise the importance weights:

$$\widetilde{w}_k^{(i)} = \frac{w_k^{*(i)}}{\sum_{j=1}^{N} w_k^{*(j)}} \tag{8}$$

■

A special case of this algorithm was introduced in 1969 by (Handschin et. al, 1969)(Handschin 1970). Many of the other algorithms proposed in the literature are later shown also to be special cases of this general (and simple) algorithm. Choice of importance function is of course crucial and one obtains poor performance when the importance function is not well chosen. This issue forms the topic of the following subsection.

C. Degeneracy of the algorithm

If Bayesian IS is interpreted as a Monte Carlo sampling method rather than as a Monte Carlo integration method, the best possible choice of importance function is of course the posterior distribution itself, $p\left(\mathbf{x}_{0:k} \middle| \mathbf{y}_{0:k}\right)$. We would ideally like to be close to this case. However, for importance functions of the form (6), the variance of the importance weights can only increase (stochastically) over time.

Proposition 1 *The unconditional variance of the importance weights, i.e. with the observations* $\mathbf{y}_{0:k}$ *being interpreted as random variables, increases over time.*

The proof of this proposition is a straightforward extension of a Kong-Liu-Wong theorem (Kong et al., 1994) to the case of an importance function of the form (6). Thus, it is impossible to avoid a degeneracy phenomenon. In practice, after a few iterations of the algorithm, all but one of the normalised importance weights are very close to zero and a large computational effort is devoted to updating trajectories whose contribution to the final estimate is almost zero.

D. Selection of the importance function

To limit degeneracy of the algorithm, a natural strategy consists of selecting the importance function which minimises the variance of the importance weights conditional upon the simulated trajectory $\mathbf{x}_{0:k-1}^{(i)}$ and the observations $\mathbf{y}_{0:k}$.

Proposition 2 $\pi(\mathbf{x}_k|\, \mathbf{x}_{0:k-1}, \mathbf{y}_{0:k}) = p(\mathbf{x}_k|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_k)$ *is the importance function which minimises the variance of the importance weight* $w_k^{*(i)}$ *conditional upon* $\mathbf{x}_{0:k-1}^{(i)}$ *and* $\mathbf{y}_{0:k}$.

Proof. Straightforward calculations yield

$$var_{\pi\left(\mathbf{x}_k|\mathbf{x}_{0:k-1}^{(i)},\mathbf{y}_{0:k}\right)}\left[w_k^{*(i)}\right] = \frac{\left(w_{k-1}^{*(i)}\right)^2}{p^2\left(\mathbf{y}_k|\, \mathbf{y}_{0:k-1}\right)}\left[\int \frac{\left(p\left(\mathbf{y}_k|\, \mathbf{x}_k\right) p\left(\mathbf{x}_k|\, \mathbf{x}_{k-1}^{(i)}\right)\right)^2}{\pi\left(\mathbf{x}_k|\, \mathbf{y}_{0:k}, \mathbf{x}_{0:k-1}^{(i)}\right)} d\mathbf{x}_k - p^2\left(\mathbf{y}_k|\, \mathbf{x}_{k-1}^{(i)}\right)\right]$$

This variance is zero for $\pi\left(\mathbf{x}_k|\, \mathbf{y}_{0:k}, \mathbf{x}_{0:k-1}^{(i)}\right) = p\left(\mathbf{x}_k|\, \mathbf{y}_k, \mathbf{x}_{k-1}^{(i)}\right)$. ■

1. Optimal importance function

The optimal importance function $p\left(\mathbf{x}_k|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_k\right)$ was introduced by (Zaritskii et al., 1975) then by (Akashi et al., 1977) for a particular case. More recently, this importance function has been used in (Chen et al., 1996)(Kong et al., 1994)(Liu et al., 1995). For this distribution, we obtain using (7) for the importance weight $w_k^{*(i)} = w_{k-1}^{*(i)} p\left(\mathbf{y}_k|\, \mathbf{x}_{k-1}^{(i)}\right)$. The optimal importance function suffers from two major drawbacks. It requires the ability to sample from $p\left(\mathbf{x}_k|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_k\right)$ and to evaluate, up to a proportionality constant, $p\left(\mathbf{y}_k|\, \mathbf{x}_{k-1}^{(i)}\right) =$

$\int p\left(\mathbf{y}_k|\,\mathbf{x}_k\right) p\left(\mathbf{x}_k|\,\mathbf{x}_{k-1}^{(i)}\right) d\mathbf{x}_k$. This integral will have no analytic form in the general case. Nevertheless, analytic evaluation is possible for the important class of models presented below, the Gaussian state space model with non-linear transition equation.

Example 3 *Nonlinear Gaussian State Space Models. Let us consider the following model:*

$$\mathbf{x}_k = f\left(\mathbf{x}_{k-1}\right) + \mathbf{v}_k,\ \mathbf{v}_k \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{v}}\right) \tag{9}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{w}_k,\ \mathbf{w}_k \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{w}}\right) \tag{10}$$

where $f : \mathbb{R}^{n_{\mathbf{x}}} \to \mathbb{R}^{n_{\mathbf{x}}}$ *is a real-valued non-linear function,* $\mathbf{C} \in \mathbb{R}^{n_y \times n_x}$ *is an observation matrix, and* $\mathbf{v}_k$ *and* $\mathbf{w}_k$ *are mutually independent i.i.d.Gaussian sequences with* $\mathbf{\Sigma}_{\mathbf{v}} > 0$ *and* $\mathbf{\Sigma}_{\mathbf{w}} > 0$*,* $\mathbf{\Sigma}_{\mathbf{v}}$ *and* $\mathbf{\Sigma}_{\mathbf{w}}$ *being assumed known. Defining*

$$\mathbf{\Sigma}^{-1} = \mathbf{\Sigma}_{\mathbf{v}}^{-1} + \mathbf{C}^t \mathbf{\Sigma}_{\mathbf{w}}^{-1} \mathbf{C} \tag{11}$$

$$\mathbf{m}_k = \mathbf{\Sigma}\left(\mathbf{\Sigma}_{\mathbf{v}}^{-1} f\left(\mathbf{x}_{k-1}\right) + \mathbf{C}^t \mathbf{\Sigma}_{\mathbf{w}}^{-1} \mathbf{y}_k\right) \tag{12}$$

one obtains

$$\mathbf{x}_k|\,\mathbf{x}_{k-1}, \mathbf{y}_k \sim \mathcal{N}\left(\mathbf{m}_k, \mathbf{\Sigma}\right) \tag{13}$$

and

$$p\left(\mathbf{y}_k|\,\mathbf{x}_{k-1}\right) \propto \exp\left(-\frac{1}{2}\left(\mathbf{y}_k - \mathbf{C}f\left(\mathbf{x}_{k-1}\right)\right)^t \left(\mathbf{\Sigma}_{\mathbf{v}} + \mathbf{C}\mathbf{\Sigma}_{\mathbf{w}}\mathbf{C}^t\right)^{-1} \left(\mathbf{y}_k - \mathbf{C}f\left(\mathbf{x}_{k-1}\right)\right)\right) \tag{14}$$

For many other models, such evaluations are impossible. We now present suboptimal methods which allow approximation of the optimal importance function. Several Monte Carlo methods have been proposed to approximate the importance function and the associated importance weight based on importance sampling (Doucet, 1997)(Doucet, 1998) and Markov chain Monte Carlo methods (Berzuini et al., 1998)(Liu et al., 1998). These iterative algorithms are computationally intensive and there is a lack of theoretical convergence results. However, these methods may be useful when non-iterative schemes fail. In fact, the general framework of SIS allows us to consider other importance functions built so as to approximate analytically the optimal importance function. The advantages of this alternative approach are that it is computationally less expensive than Monte Carlo methods and that the standard convergence results for Bayesian importance sampling are still valid. There is no general method to build suboptimal importance functions and it is necessary to build these on a case by case basis, dependent on the model studied. To this end, it is possible to base these developments on previous work in suboptimal filtering (Anderson et al., 1979)(West et al., 1997), and this is considered in the next subsection.

2. Importance distribution obtained by local linearisation

A simple choice selects as the importance function $\pi(\mathbf{x}_k|\, \mathbf{x}_{k-1}, \mathbf{y}_k)$ a parametric distribution $\pi(\mathbf{x}_k|\, \boldsymbol{\theta}(\mathbf{x}_{k-1}, \mathbf{y}_k))$, with finite-dimensional parameter $\boldsymbol{\theta}$ ($\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{n_{\boldsymbol{\theta}}}$) determined by $\mathbf{x}_{k-1}$ and $\mathbf{y}_k$, $\boldsymbol{\theta} : \mathbb{R}^{n_{\mathbf{x}}} \times \mathbb{R}^{n_{\mathbf{y}}} \rightarrow \boldsymbol{\Theta}$ being a deterministic mapping. Many strategies are possible based upon this idea. To illustrate such methods, we present here two novel schemes that result in a Gaussian importance function whose parameters are evaluated using local linearisations, i.e. which are dependent on the simulated trajectory $i = 1, ..., N$. Such an approach seems to be a very promising way of proceeding with many models, where linearisations are readily and cheaply available. In the auxiliary variables framework of (Pitt and Shephard, 1999), related 'suboptimal' importance distributions are proposed to sample efficiently from a finite mixture distribution approximating the filtering distribution. We follow here a different approach in which the filtering distribution is approximated directly without resort to auxiliary indicator variables.

Local linearisation of the state space model We propose to linearise the model locally in a similar way to the Extended Kalman Filter. However, in our case, this linearisation is performed with the aim of obtaining an importance function and the algorithm obtained still converges asymptotically towards the required filtering distribution under the usual assumptions for importance functions.

Example 4 *Let us consider the following model*

$$\begin{aligned} \mathbf{x}_k &= f(\mathbf{x}_{k-1}) + \mathbf{v}_k,\ \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}_{n_\mathbf{v}\times 1}, \mathbf{\Sigma}_\mathbf{v}) && (15) \\ \mathbf{y}_k &= g(\mathbf{x}_k) + \mathbf{w}_k,\ \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}_{n_\mathbf{w}\times 1}, \mathbf{\Sigma}_\mathbf{w}) && (16) \end{aligned}$$

where $f : \mathbb{R}^{n_\mathbf{x}} \to \mathbb{R}^{n_\mathbf{x}}$*,* $g : \mathbb{R}^{n_\mathbf{x}} \to \mathbb{R}^{n_\mathbf{y}}$ *are differentiable,* $\mathbf{v}_k$ *and* $\mathbf{w}_k$ *are two mutually independent i.i.d. sequences with* $\mathbf{\Sigma}_\mathbf{v} > 0$ *and* $\mathbf{\Sigma}_\mathbf{w} > 0$*. Performing an approximation up to first order of the observation equation (Anderson et al., 1979), we get*

$$\begin{aligned} \mathbf{y}_k &= g(\mathbf{x}_k) + \mathbf{w}_k \\ &\simeq g(f(\mathbf{x}_{k-1})) + \left.\frac{\partial g(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k = f(\mathbf{x}_{k-1})} (\mathbf{x}_k - f(\mathbf{x}_{k-1})) + \mathbf{w}_k && (17) \end{aligned}$$

We have now defined a new model with a similar evolution equation to (15) but with a linear Gaussian observation equation (17), obtained by linearising $g(\mathbf{x}_k)$ *in* $f(\mathbf{x}_{k-1})$*. This model is not Markovian as (17) depends on* $\mathbf{x}_{k-1}$*. However, it is of the form (9)-(10) and one can perform similar calculations to obtain a Gaussian importance function* $\pi(\mathbf{x}_k|\,\mathbf{x}_{k-1}, \mathbf{y}_k) \sim \mathcal{N}(\mathbf{m}_k, \mathbf{\Sigma}_k)$ *with mean* $\mathbf{m}_k$ *and covariance* $\mathbf{\Sigma}_k$ *evaluated for each trajectory* $i = 1, ..., N$ *using the following formula:*

$$\mathbf{\Sigma}_k^{-1} = \mathbf{\Sigma}_\mathbf{v}^{-1} + \left[\left.\frac{\partial g(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k = f(\mathbf{x}_{k-1})}\right]^t \mathbf{\Sigma}_\mathbf{w}^{-1} \left.\frac{\partial g(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k = f(\mathbf{x}_{k-1})} \tag{18}$$

$$\begin{aligned} \mathbf{m}_k &= \mathbf{\Sigma}_k \left(\mathbf{\Sigma}_\mathbf{v}^{-1} f(\mathbf{x}_{k-1}) + \left[\left.\frac{\partial g(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k = f(\mathbf{x}_{k-1})}\right]^t \mathbf{\Sigma}_\mathbf{w}^{-1} \times \right. && (19) \\ &\quad \left. \times \left(\mathbf{y}_k - g(f(\mathbf{x}_{k-1})) + \left.\frac{\partial g(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k = f(\mathbf{x}_{k-1})} f(\mathbf{x}_{k-1}) \right) \right) && (20) \end{aligned}$$

The associated importance weight is evaluated using (7).

Local linearisation of the optimal importance function We assume here that $l(\mathbf{x}_k) \triangleq \log p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k)$ is twice differentiable wrt $\mathbf{x}_k$ on $\mathbb{R}^{n_\mathbf{x}}$. We define:

$$l'(\mathbf{x}) \triangleq \left.\frac{\partial l(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\mathbf{x}_k=\mathbf{x}} \tag{21}$$

$$l''(\mathbf{x}) \triangleq \left.\frac{\partial^2 l(\mathbf{x}_k)}{\partial \mathbf{x}_k \partial \mathbf{x}_k^t}\right|_{\mathbf{x}_k=\mathbf{x}} \tag{22}$$

Using a second order Taylor expansion in $\mathbf{x}$, we get :

$$l(\mathbf{x}_k) \simeq l(\mathbf{x}) + \left[l'(\mathbf{x})\right]^t (\mathbf{x}_k - \mathbf{x}) + \frac{1}{2} (\mathbf{x}_k - \mathbf{x})^t l''(\mathbf{x}) (\mathbf{x}_k - \mathbf{x}) \tag{23}$$

The point $\mathbf{x}$ where we perform the expansion is arbitrary (but determined by a deterministic mapping of $\mathbf{x}_{k-1}$ and $\mathbf{y}_k$). Under the additional assumption that $l''(\mathbf{x})$ is negative definite, which is true if $l(\mathbf{x}_k)$ is concave, then setting

$$\boldsymbol{\Sigma}(\mathbf{x}) = -l''(\mathbf{x})^{-1} \tag{24}$$

$$\mathbf{m}(\mathbf{x}) = \boldsymbol{\Sigma}(\mathbf{x}) l'(\mathbf{x}) \tag{25}$$

yields

$$\begin{aligned} & \left[l'(\mathbf{x})\right]^t (\mathbf{x}_k - \mathbf{x}) + \frac{1}{2} (\mathbf{x}_k - \mathbf{x})^t l''(\mathbf{x}) (\mathbf{x}_k - \mathbf{x}) \\ = \;& C - \frac{1}{2} (\mathbf{x}_k - \mathbf{x} - \mathbf{m}(\mathbf{x}))^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}) (\mathbf{x}_k - \mathbf{x} - \mathbf{m}(\mathbf{x})) \end{aligned} \tag{26}$$

This suggests adoption of the following importance function:

$$\pi(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k) = \mathcal{N}(\mathbf{m}(\mathbf{x}) + \mathbf{x}, \boldsymbol{\Sigma}(\mathbf{x})) \tag{27}$$

If $p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k)$ is unimodal, it is judicious to adopt $\mathbf{x}$ as the mode of $p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k)$, thus $\mathbf{m}(\mathbf{x}) = \mathbf{0}_{n_\mathbf{x} \times 1}$. The associated importance weight is evaluated using (7).

Example 5 *Linear Gaussian Dynamic/Observations according to a distribution from the exponential family. We assume that the evolution equation satisfies:*

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{v}_k \textit{ where } \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}_{n_\mathbf{v} \times 1}, \boldsymbol{\Sigma}_\mathbf{v}) \tag{28}$$

where $\mathbf{\Sigma_v} > 0$ *and the observations are distributed according to a distribution from the exponential family, i.e.*

$$p(\mathbf{y}_k|\,\mathbf{x}_k) = \exp\left(\mathbf{y}_k^t\mathbf{C}\mathbf{x}_k - b(\mathbf{C}\mathbf{x}_k) + c(\mathbf{y}_k)\right) \tag{29}$$

where $\mathbf{C}$ *is a real* $n_{\mathbf{y}} \times n_{\mathbf{x}}$ *matrix,* $b : \mathbb{R}^{n_{\mathbf{y}}} \to \mathbb{R}$ *and* $c : \mathbb{R}^{n_{\mathbf{y}}} \to \mathbb{R}$. *These models have numerous applications and allow consideration of Poisson or binomial observations, see for example (West et al., 1997). We have*

$$l(\mathbf{x}_k) = C + \mathbf{y}_k^t\mathbf{C}\mathbf{x}_k - b(\mathbf{C}\mathbf{x}_k) - \frac{1}{2}(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1})^t\,\mathbf{\Sigma_v}^{-1}(\mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1}) \tag{30}$$

This yields

$$\begin{aligned} l''(\mathbf{x}) &= -\left.\frac{\partial^2 b(\mathbf{C}\mathbf{x}_k)}{\partial\mathbf{x}_k\partial\mathbf{x}_k^t}\right|_{\mathbf{x}_k=\mathbf{x}} - \mathbf{\Sigma_v}^{-1} \\ &= -b''(\mathbf{x}) - \mathbf{\Sigma_v}^{-1} \end{aligned} \tag{31}$$

but $b''(\mathbf{x})$ *is the covariance matrix of* $\mathbf{y}_k$ *for* $\mathbf{x}_k = \mathbf{x}$, *thus* $l''(\mathbf{x})$ *is definite negative. One can determine the mode* $\mathbf{x} = \mathbf{x}^*$ *of this distribution by applying an iterative Newton-Raphson method initialised with* $\mathbf{x}_{(0)} = \mathbf{x}_{k-1}$, *which satisfies at iteration* j*:*

$$\mathbf{x}_{(j+1)} = \mathbf{x}_{(j)} - \left[l''\left(\mathbf{x}_{(j)}\right)\right]^{-1} l'\left(\mathbf{x}_{(j)}\right) \tag{32}$$

We now present two simpler importance functions which lead to algorithms which previously appeared in the literature.

3. Prior importance function

A simple choice uses the prior distribution of the hidden Markov model as importance function. This is the choice made by (Handschin et. al, 1969)(Handschin 1970) in their seminal work. This is one of the methods recently proposed in (Tanizaki et al., 1998). In this case, we have $\pi(\mathbf{x}_k|\,\mathbf{x}_{0:k-1}, \mathbf{y}_{0:k}) = p(\mathbf{x}_k|\,\mathbf{x}_{k-1})$ and $w_k^{*(i)} = w_{k-1}^{*(i)} p\left(\mathbf{y}_k|\,\mathbf{x}_k^{(i)}\right)$. The method is often inefficient in simulations as the state space is explored without any knowledge of the observations. It is especially sensitive to outliers. However, it does have the advantage that the importance weights are easily evaluated. Use of the prior importance function is closely related to the Bootstrap filter method of (Gordon et al., 1993), see Section III..

4. Fixed importance function

An even simpler choice fixes an importance function independently of the simulated trajectories and of the observations. In this case, we have $\pi\left(\mathbf{x}_k \mid \mathbf{x}_{0:k-1}, \mathbf{y}_{0:k}\right) = \pi\left(\mathbf{x}_k\right)$ and

$$w_k^{*(i)} = w_{k-1}^{*(i)} p\left(\mathbf{y}_k \mid \mathbf{x}_k^{(i)}\right) p\left(\mathbf{x}_k^{(i)} \middle| \mathbf{x}_{k-1}^{(i)}\right) / \pi\left(\mathbf{x}_k^{(i)}\right) \tag{33}$$

This is the importance function adopted by (Tanizaki, 1993)(Tanizaki, 1994) who present this method as a stochastic alternative to the numerical integration method of (Kitagawa, 1987). The results obtained are rather poor as neither the dynamic of the model nor the observations are taken into account and leads in most cases to unbounded (unnormalised) importance weights which will give poor results (Geweke, 1989).

III. Resampling

As has previously been illustrated, the degeneracy of the SIS algorithm is unavoidable. The basic idea of resampling methods is to eliminate trajectories which have small normalised importance weights and to concentrate upon trajectories with large weights. A suitable measure of degeneracy of the algorithm is the effective sample size N_{eff} introduced in (Kong et al., 1994)(Liu, 1996) and defined as:

$$N_{eff} = \frac{N}{1 + var_{\pi(\cdot|\mathbf{y}_{0:k})}\left(w^*\left(\mathbf{x}_{0:k}\right)\right)} = \frac{N}{\mathbb{E}_{\pi(\cdot|\mathbf{y}_{0:k})}\left[\left(w^*\left(\mathbf{x}_{0:k}\right)\right)^2\right]} \leq N \tag{34}$$

One cannot evaluate N_{eff} exactly but, an estimate $\widehat{N_{eff}}$ of N_{eff} is given by:

$$\widehat{N_{eff}} = \frac{N}{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{w}_k^{*(i)}\right)^2} = \frac{1}{\sum_{i=1}^{N}\left(\widetilde{w}_k^{(i)}\right)^2} \tag{35}$$

When $\widehat{N_{eff}}$ is below a fixed threshold N_{thres}, the SIR resampling procedure is used (Rubin, 1988). Note that it is possible to implement the SIR procedure exactly in $O(N)$ operations by using a classical algorithm (Ripley, 1987 p. 96) and (Carpenter et al., 1997)(Doucet, 1997)(Doucet, 1998)(Pitt et al., 1999). Other resampling procedures which reduce the MC variation, such as stratified sampling (Carpenter et al., 1997) and residual resampling (Liu et al., 1998), may be applied as an alternative to SIR.

An appropriate algorithm based on the SIR scheme proceeds as follows at time k.

SIS/Resampling Monte Carlo filter

1. Importance sampling

   - For $i = 1, ..., N$, sample $\widetilde{\mathbf{x}}_k^{(i)} \sim \pi(\mathbf{x}_k | \mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}_{0:k})$ and $\widetilde{\mathbf{x}}_{0:k}^{(i)} \triangleq \left(\mathbf{x}_{0:k-1}^{(i)}, \widetilde{\mathbf{x}}_k^{(i)}\right)$.
   - For $i = 1, ..., N$, evaluate the importance weights up to a normalising constant:

$$w_k^{*(i)} = w_{k-1}^{*(i)} \frac{p\left(\mathbf{y}_k | \widetilde{\mathbf{x}}_k^{(i)}\right) p\left(\widetilde{\mathbf{x}}_k^{(i)} \middle| \widetilde{\mathbf{x}}_{k-1}^{(i)}\right)}{\pi\left(\widetilde{\mathbf{x}}_k^{(i)} \middle| \widetilde{\mathbf{x}}_{0:k-1}^{(i)}, \mathbf{y}_{0:k}\right)} \tag{36}$$

   - For $i = 1, ..., N$, normalise the importance weights:

$$\widetilde{w}_k^{(i)} = \frac{w_k^{*(i)}}{\sum_{j=1}^{N} w_k^{*(j)}} \tag{37}$$

   - Evaluate $\widehat{N_{eff}}$ using (35).

2. Resampling

   If $\widehat{N_{eff}} \geq N_{thres}$

   - $\mathbf{x}_{0:k}^{(i)} = \widetilde{\mathbf{x}}_{0:k}^{(i)}$ for $i = 1, ..., N$.

   otherwise

   - For $i = 1, ..., N$, sample an index $j(i)$ distributed according to the discrete distribution with N elements satisfying $\Pr\{j(i) = l\} = \widetilde{w}_k^{(l)}$ for $l = 1, ..., N$.

- For $i = 1, ..., N$, $\mathbf{x}_{0:k}^{(i)} = \widetilde{\mathbf{x}}_{0:k}^{j(i)}$ and $w_k^{(i)} = \frac{1}{N}$.

If $\widehat{N_{eff}} \geq N_{thres}$, the algorithm presented in Subsection B. is thus not modified and if $\widehat{N_{eff}} < N_{thres}$ the SIR algorithm is applied and one obtains

$$\widehat{P}\left(d\mathbf{x}_{0:k} \middle| \mathbf{y}_{0:k}\right) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\mathbf{x}_{0:k}^{(i)}}\left(d\mathbf{x}_{0:k}\right) \tag{38}$$

Resampling procedures decrease algorithmically the degeneracy problem but introduce practical and theoretical problems. From a theoretical point of view, after one resampling step, the simulated trajectories are no longer statistically independent and so we lose the simple convergence results given previously. Recently, (Berzuini et al., 1998) have however established a central limit theorem for the estimate of $I(f_k)$ obtained when the SIR procedure is applied at each iteration. From a practical point of view, the resampling scheme limits the opportunity to parallelise since all the particles must be combined, although the IS steps can still be realized in parallel. Moreover the trajectories $\left\{\widetilde{\mathbf{x}}_{0:k}^{(i)}, i = 1, ..., N\right\}$ which have high importance weights $\widetilde{w}_k^{(i)}$ are statistically selected many times. In (38), numerous trajectories $\mathbf{x}_{0:k}^{(i_1)}$ and $\mathbf{x}_{0:k}^{(i_2)}$ are in fact equal for $i_1 \neq i_2 \in [1, \ldots, N]$. There is thus a loss of "diversity". Various heuristic methods have been proposed to solve this problem (Gordon et al., 1993)(Higuchi, 1997).

IV. Rao-Blackwellisation for Sequential Importance Sampling

In this section we describe variance reduction methods which are designed to make the most of any structure within the model studied. Numerous methods have been developed for reducing the variance of MC estimates including antithetic sampling (Handschin et. al, 1969)(Handschin 1970) and control variates (Akashi et al., 1975)(Handschin 1970). We apply here the Rao-Blackwellisation method, see (Casella et al. 1996) for a general reference on the topic. In a sequential framework, (MacEachern et al. 1998) have applied similar ideas for Dirichlet process models and (Kong et al. 1994)(Liu et al. 1998) have used Rao-Blackwellisation for fixed parameter estimation. We focus on its application to dynamic models. We show how it is possible to successfully apply this method to an important class of state space model and obtain hybrid filters where a part of the calculations is realised analytically and the other part using MC methods.

The following method is useful for cases when one can partition the state $\mathbf{x}_k$ as $\left(\mathbf{x}_k^1, \mathbf{x}_k^2\right)$and analytically marginalize one component of the partition, say $\mathbf{x}_k^2$. For instance, as demonstrated in example 6, if one component of the partition is a conditionally linear Gaussian state-space model then all the integrations can be performed analytically on-line using the Kalman filter. Let us define $\mathbf{x}_{0:n}^j \triangleq \left(\mathbf{x}_0^j, \ldots, \mathbf{x}_n^j\right)$. We can rewrite the posterior expectation $I\left(f_n\right)$ in terms of marginal quantities:

$$
\begin{aligned}
I\left(f_n\right) &= \frac{\int \left[\int f_n\left(\mathbf{x}_{0:n}^1, \mathbf{x}_{0:n}^2\right) p\left(\mathbf{y}_{0:n} \middle| \mathbf{x}_{0:n}^1, \mathbf{x}_{0:n}^2\right) p\left(\mathbf{x}_{0:n}^2 \middle| \mathbf{x}_{0:n}^1\right) d\mathbf{x}_{0:n}^2\right] p\left(\mathbf{x}_{0:n}^1\right) d\mathbf{x}_{0:n}^1}{\int \left[\int p\left(\mathbf{y}_{0:n} \middle| \mathbf{x}_{0:n}^1, \mathbf{x}_{0:n}^2\right) p\left(\mathbf{x}_{0:n}^2 \middle| \mathbf{x}_{0:n}^1\right) d\mathbf{x}_{0:n}^2\right] p\left(\mathbf{x}_{0:n}^1\right) d\mathbf{x}_{0:n}^1} \\
&= \frac{\int g(\mathbf{x}_{0:n}^1) p\left(\mathbf{x}_{0:n}^1\right) d\mathbf{x}_{0:n}^1}{\int p\left(\mathbf{y}_{0:n} \middle| \mathbf{x}_{0:n}^1\right) p\left(\mathbf{x}_{0:n}^1\right) d\mathbf{x}_{0:n}^1}
\end{aligned}
$$

where

$$
g(\mathbf{x}_{0:n}^1) \triangleq \int f_n\left(\mathbf{x}_{0:n}^1, \mathbf{x}_{0:n}^2\right) p\left(\mathbf{y}_{0:n} \middle| \mathbf{x}_{0:n}^1, \mathbf{x}_{0:n}^2\right) p\left(\mathbf{x}_{0:n}^2 \middle| \mathbf{x}_{0:n}^1\right) d\mathbf{x}_{0:n}^2 \tag{39}
$$

Under the assumption that, conditional upon a realisation of $\mathbf{x}_{0:n}^1$, $g(\mathbf{x}_{0:n}^1)$ and $p\left(\mathbf{y}_{0:n} \middle| \mathbf{x}_{0:n}^1\right)$ can be evaluated analytically, two estimates of $I\left(f_n\right)$ based on IS are possible. The first "classical" one is obtained using as importance distribution $\pi\left(\mathbf{x}_{0:n}^1, \mathbf{x}_{0:n}^2 \middle| \mathbf{y}_{0:n}\right)$:

$$
\widehat{I_N}\left(f_n\right) = \frac{\sum_{i=1}^N f_n\left(\mathbf{x}_{0:n}^{1,(i)}, \mathbf{x}_{0:n}^{2,(i)}\right) w^*\left(\mathbf{x}_{0:n}^{1,(i)}, \mathbf{x}_{0:n}^{2,(i)}\right)}{\sum_{i=1}^N w^*\left(\mathbf{x}_{0:n}^{1,(i)}, \mathbf{x}_{0:n}^{2,(i)}\right)} \tag{40}
$$

where $w^*\left(\mathbf{x}_{0:n}^{1,(i)}, \mathbf{x}_{0:n}^{2,(i)}\right) \propto p\left(\mathbf{x}_{0:n}^{1,(i)}, \mathbf{x}_{0:n}^{2,(i)} \middle| \mathbf{y}_{0:n}\right) / \pi\left(\mathbf{x}_{0:n}^{1,(i)}, \mathbf{x}_{0:n}^{2,(i)} \middle| \mathbf{y}_{0:n}\right)$. The second "Rao-Blackwellised" estimate is obtained by analytically integrating out $\mathbf{x}_{0:n}^2$ and using as importance distribution $\pi\left(\mathbf{x}_{0:n}^1 \middle| \mathbf{y}_{0:n}\right) = \int \pi\left(\mathbf{x}_{0:n}^1, \mathbf{x}_{0:n}^2 \middle| \mathbf{y}_{0:n}\right) d\mathbf{x}_{0:n}^2$. The new estimate is given by:

$$
\widehat{\widehat{I_N}}\left(f_n\right) = \frac{\sum_{i=1}^N g\left(\mathbf{x}_{0:n}^{1,(i)}\right) w^*\left(\mathbf{x}_{0:n}^{1,(i)}\right)}{\sum_{i=1}^N w^*\left(\mathbf{x}_{0:n}^{1,(i)}\right)} \tag{41}
$$

where $w^*\left(\mathbf{x}_{0:n}^{1,(i)}\right) \propto p\left(\mathbf{x}_{0:n}^{1,(i)}\middle|\,\mathbf{y}_{0:n}\right)/\pi\left(\mathbf{x}_{0:n}^{1,(i)}\middle|\,\mathbf{y}_{0:n}\right)$. Using the decomposition of the variance, it is straightforward to show that the variances of the importance weights obtained by Rao-Blackwellisation are smaller than those obtained using a direct Monte Carlo method (40), see for example (Doucet 1997)(Doucet 1998)(MacEachern et al. 1998). We can use this method to estimate $I(f_n)$ and marginal quantities such as $p\left(\mathbf{x}_{0:n}^1\middle|\,\mathbf{y}_{0:n}\right)$.

One has to be cautious when applying the MC methods developed in the previous sections to the marginal state space $\mathbf{x}_k^1$. Indeed, even if the observations $\mathbf{y}_{0:n}$ are independent conditional upon $\left(\mathbf{x}_{0:n}^1, \mathbf{x}_{0:n}^2\right)$, they are generally no longer independent conditional upon the single process $\mathbf{x}_{0:n}^1$. The required modifications are, however, straightforward. For example, we obtain for the optimal importance function $p\left(\mathbf{x}_k^1\middle|\,\mathbf{y}_{0:k}, \mathbf{x}_{0:k-1}^1\right)$ and its associated importance weight $p\left(\mathbf{y}_k\middle|\,\mathbf{y}_{0:k-1}, \mathbf{x}_{0:k-1}^1\right)$. We now present two important applications of this general method.

Example 6 *Conditionally linear Gaussian state space model*

Let us consider the following model

$$p\left(\mathbf{x}_k^1\middle|\,\mathbf{x}_{k-1}^1\right) \tag{42}$$

$$\mathbf{x}_k^2 = \mathbf{A}_k\left(\mathbf{x}_k^1\right)\mathbf{x}_{k-1}^2 + \mathbf{B}_k\left(\mathbf{x}_k^1\right)\mathbf{v}_k \tag{43}$$

$$\mathbf{y}_k = \mathbf{C}_k\left(\mathbf{x}_k^1\right)\mathbf{x}_k^2 + \mathbf{D}_k\left(\mathbf{x}_k^1\right)\mathbf{w}_k \tag{44}$$

where $\mathbf{x}_k^1$ is a Markov process, $\mathbf{v}_k \sim \mathcal{N}\left(\mathbf{0}_{n_\mathbf{v}\times 1}, \mathbf{I}_{n_\mathbf{v}}\right)$ and $\mathbf{w}_k \sim \mathcal{N}\left(\mathbf{0}_{n_\mathbf{w}\times 1}, \mathbf{I}_{n_\mathbf{w}}\right)$. One wants to estimate $p\left(\mathbf{x}_{0:n}^1\middle|\,\mathbf{y}_{0:n}\right)$, $\mathbb{E}\left(f\left(\mathbf{x}_n^1\right)\middle|\,\mathbf{y}_{0:n}\right)$, $\mathbb{E}\left(\mathbf{x}_n^2\middle|\,\mathbf{y}_{0:n}\right)$ and $\mathbb{E}\left(\mathbf{x}_n^2\left(\mathbf{x}_n^2\right)^t\middle|\,\mathbf{y}_{0:n}\right)$. It is possible to use a MC filter based on Rao-Blackwellisation. Indeed, conditional upon $\mathbf{x}_{0:n}^1$, $\mathbf{x}_{0:n}^2$ is a linear Gaussian state space model and the integrations required by the Rao-Blackwellisation method can be realized using the Kalman filter.

Akashi and Kumamoto (Akashi et al., 1977)(Tugnait, 1982) introduced this algorithm under the name of RSA (Random Sampling Algorithm) in the particular case where $\mathbf{x}_k^1$ is a homogeneous scalar finite state-space Markov chain. In this case, they adopted the optimal importance function $p\left(\left.\mathbf{x}_k^1\right|\mathbf{y}_{0:k},\mathbf{x}_{0:k-1}^1\right)$. Indeed, it is possible to sample from this discrete distribution and to evaluate the importance weight $p\left(\left.\mathbf{y}_k\right|\mathbf{y}_{0:k},\mathbf{x}_{0:k-1}^1\right)$ using the Kalman filter (Akashi et al., 1977). Similar developments for this special case have also been proposed by (Svetnik, 1986)(Billio et al., 1998)(Liu et al., 1998). The algorithm for blind deconvolution proposed by (Liu et al., 1995) is also a particular case of this method where $\mathbf{x}_k^2 = \mathbf{h}$ is a time-invariant channel of Gaussian prior distribution. Using the Rao-Blackwellisation method in this framework is particularly attractive as, while $\mathbf{x}_k$ has some continuous components, we restrict ourselves to the exploration of a discrete state space.

Example 7 *Finite State-Space HMM*

Let us consider the following model

$$p\left(\left.\mathbf{x}_k^1\right|\mathbf{x}_{k-1}^1\right) \tag{45}$$

$$p\left(\left.\mathbf{x}_k^2\right|\mathbf{x}_k^1,\mathbf{x}_{k-1}^2\right) \tag{46}$$

$$p\left(\left.\mathbf{y}_k\right|\mathbf{x}_k^1,\mathbf{x}_k^2\right) \tag{47}$$

where $\mathbf{x}_k^1$ is a Markov process and $\mathbf{x}_k^2$ is a finite state-space Markov chain whose parameters at time k depend on $\mathbf{x}_k^1$. We want to estimate $p\left(\left.\mathbf{x}_{0:n}^1\right|\mathbf{y}_{0:n}\right)$, $\mathbb{E}\left(\left.f\left(\mathbf{x}_n^1\right)\right|\mathbf{y}_{0:n}\right)$ and $\mathbb{E}\left(\left.f\left(\mathbf{x}_n^2\right)\right|\mathbf{y}_{0:n}\right)$. It is possible to use a "Rao-Blackwellised" MC filter. Indeed, conditional upon $\mathbf{x}_{0:n}^1$, $\mathbf{x}_{0:n}^2$ is a finite state-space Markov chain of known parameters and thus the integrations required by the Rao-Blackwellisation method can be done analytically (Anderson et al., 1979).

V. Prediction, smoothing and likelihood

The estimate of the joint distribution $p\left(\mathbf{x}_{0:k}\middle|\,\mathbf{y}_{0:k}\right)$ based on SIS, in practice coupled with a resampling procedure to limit the degeneracy, is at any time k of the following form:

$$\widehat{P}\left(d\mathbf{x}_{0:k}\middle|\,\mathbf{y}_{0:k}\right)=\sum_{i=1}^{N}\widetilde{w}_{k}^{(i)}\delta_{\mathbf{x}_{0:k}^{(i)}}\left(d\mathbf{x}_{0:k}\right) \tag{48}$$

We show here how it is possible to obtain based on this distribution some approximations of the prediction and smoothing distributions as well as the likelihood.

A. Prediction

Based on the approximation of the filtering distribution $\widehat{P}\left(d\mathbf{x}_{k}\middle|\,\mathbf{y}_{0:k}\right)$, we want to estimate the p step-ahead prediction distribution, $p\geq 2\in\mathbb{N}^{*}$, given by:

$$p\left(\mathbf{x}_{k+p}\middle|\,\mathbf{y}_{0:k}\right)=\int p\left(\mathbf{x}_{k}\middle|\,\mathbf{y}_{0:k}\right)\left[\prod_{j=k+1}^{k+p}p\left(\mathbf{x}_{j}\middle|\,\mathbf{x}_{j-1}\right)\right]d\mathbf{x}_{k:k+p-1} \tag{49}$$

Replacing $p\left(\mathbf{x}_{k}\middle|\,\mathbf{y}_{0:k}\right)$ in (49) by its approximation obtained from (48), we obtain:

$$\sum_{i=1}^{N}\widetilde{w}_{k}^{(i)}\int p\left(\mathbf{x}_{k+1}\middle|\,\mathbf{x}_{k}^{(i)}\right)\prod_{j=k+2}^{k+p}p\left(\mathbf{x}_{j}\middle|\,\mathbf{x}_{j-1}\right)d\mathbf{x}_{k+1:k+p-1} \tag{50}$$

To evaluate these integrals, it is sufficient to extend the trajectories $\mathbf{x}_{0:k}^{(i)}$ using the evolution equation.

---

p step-ahead prediction

- For $j=1$ to p
  - For $i=1,...,N$, sample $\mathbf{x}_{k+j}^{(i)}\sim p\left(\mathbf{x}_{k+j}\middle|\,\mathbf{x}_{k+j-1}^{(i)}\right)$ and $\mathbf{x}_{0:k+j}^{(i)}\triangleq\left(\mathbf{x}_{0:k+j-1}^{(i)},\mathbf{x}_{k+j}^{(i)}\right)$.

We obtain random samples $\left\{\mathbf{x}_{0:k+p}^{(i)};i=1,...,N\right\}$. An estimate of $\widehat{P}\left(d\mathbf{x}_{0:k+p}\middle|\,\mathbf{y}_{0:k}\right)$ is given by

$$\widehat{P}\left(d\mathbf{x}_{0:k+p}\middle|\,\mathbf{y}_{0:k}\right)=\sum_{i=1}^{N}\widetilde{w}_{k}^{(i)}\delta_{\mathbf{x}_{0:k+p}^{(i)}}\left(d\mathbf{x}_{0:k+p}\right)$$

Thus

$$\widehat{P}\left(d\mathbf{x}_{k+p}\right|\mathbf{y}_{0:k}) = \sum_{i=1}^{N} \widetilde{w}_k^{(i)} \delta_{\mathbf{x}_{k+p}^{(i)}}\left(d\mathbf{x}_{k+p}\right) \tag{51}$$

B. Fixed-Lag smoothing

We want to estimate the fixed-lag smoothing distribution $p\left(\mathbf{x}_k\right|\mathbf{y}_{0:k+p})$, $p \in \mathbb{N}^*$ being the length of the lag. At time $k+p$, the MC filter yields the following approximation of $p\left(\mathbf{x}_{0:k+p}\right|\mathbf{y}_{0:k+p})$:

$$\widehat{P}\left(d\mathbf{x}_{0:k+p}\right|\mathbf{y}_{0:k+p}) = \sum_{i=1}^{N} \widetilde{w}_{k+p}^{(i)} \delta_{\mathbf{x}_{0:k+p}^{(i)}}\left(d\mathbf{x}_{0:k+p}\right) \tag{52}$$

By marginalising, we obtain an estimate of the fixed-lag smoothing distribution:

$$\widehat{P}\left(d\mathbf{x}_k\right|\mathbf{y}_{0:k+p}) = \sum_{i=1}^{N} \widetilde{w}_{k+p}^{(i)} \delta_{\mathbf{x}_k^{(i)}}\left(d\mathbf{x}_k\right) \tag{53}$$

When p is high, such an approximation will generally perform poorly.

C. Fixed-interval smoothing

Given $\mathbf{y}_{0:n}$, we want to estimate $p\left(\mathbf{x}_k\right|\mathbf{y}_{0:n})$ for any $k = 0, ..., n$. At time n, the filtering algorithm yields the following approximation of $p\left(\mathbf{x}_{0:n}\right|\mathbf{y}_{0:n})$:

$$\widehat{P}\left(d\mathbf{x}_{0:n}\right|\mathbf{y}_{0:n}) = \sum_{i=1}^{N} \widetilde{w}_n^{(i)} \delta_{\mathbf{x}_{0:n}^{(i)}}\left(d\mathbf{x}_{0:n}\right) \tag{54}$$

Thus one can theoretically obtain $p\left(\mathbf{x}_k\right|\mathbf{y}_{0:n})$ for any k by marginalising this distribution. Practically, this method cannot be used as soon as $(n-k)$ is significant as the degeneracy problem requires use of a resampling algorithm. At time n, the simulated trajectories $\left\{\mathbf{x}_{0:n}^{(i)}; i = 1, ..., N\right\}$ have been usually resampled many times: there are thus only a few distinct trajectories at times k for $k \ll n$ and the above approximation of $p\left(\mathbf{x}_k\right|\mathbf{y}_{0:n})$ is bad. This problem is even more severe for the bootstrap filter where one resamples at each time instant.

It is necessary to develop an alternative algorithm. We propose an original algorithm to solve this problem. This algorithm is based on the following formula (Kitagawa, 1987):

$$p\left(\mathbf{x}_k \middle| \mathbf{y}_{0:n}\right) = p\left(\mathbf{x}_k \middle| \mathbf{y}_{0:k}\right) \int \frac{p\left(\mathbf{x}_{k+1} \middle| \mathbf{y}_{0:n}\right) p\left(\mathbf{x}_{k+1} \middle| \mathbf{x}_k\right)}{p\left(\mathbf{x}_{k+1} \middle| \mathbf{y}_{0:k}\right)} d\mathbf{x}_{k+1} \tag{55}$$

We seek here an approximation of the fixed-interval smoothing distribution with the following form:

$$\widehat{P}\left(d\mathbf{x}_k \middle| \mathbf{y}_{0:n}\right) \triangleq \sum_{i=1}^{N} \widetilde{w}_{k|n}^{(i)} \delta_{\mathbf{x}_k^{(i)}}\left(d\mathbf{x}_k\right) \tag{56}$$

i.e. $\widehat{P}\left(d\mathbf{x}_k \middle| \mathbf{y}_{0:n}\right)$ has the same support $\left\{\mathbf{x}_k^{(i)}; i = 1, \ldots, N\right\}$ as the filtering distribution $\widehat{P}\left(d\mathbf{x}_k \middle| \mathbf{y}_{0:k}\right)$ but the weights are different. An algorithm to obtain these weights $\left\{\widetilde{w}_{k|n}^{(i)}; i = 1, \ldots, N\right\}$ is the following.

Fixed-interval smoothing

1. Initialisation at time $k = n$.

   - For $i = 1, ..., N$, $\widetilde{w}_{n|n}^{(i)} = \widetilde{w}_n^{(i)}$.

2. For $k = n - 1, ..., 0$.

   - For $i = 1, ..., N$, evaluate the importance weight

$$\widetilde{w}_{k|n}^{(i)} = \sum_{j=1}^{N} \widetilde{w}_{k+1|n}^{(j)} \frac{\widetilde{w}_k^{(i)} p\left(\mathbf{x}_{k+1}^{(j)} \middle| \mathbf{x}_k^{(i)}\right)}{\left[\sum_{l=1}^{N} \widetilde{w}_k^{(l)} p\left(\mathbf{x}_{k+1}^{(j)} \middle| \mathbf{x}_k^{(l)}\right)\right]} \tag{57}$$

■

This algorithm is obtained by the following argument. Replacing $p\left(\mathbf{x}_{k+1} \middle| \mathbf{y}_{0:n}\right)$ by its approximation (56) yields

$$\int \frac{p\left(\mathbf{x}_{k+1} \middle| \mathbf{y}_{0:n}\right) p\left(\mathbf{x}_{k+1} \middle| \mathbf{x}_k\right)}{p\left(\mathbf{x}_{k+1} \middle| \mathbf{y}_{0:k}\right)} d\mathbf{x}_{k+1} \simeq \sum_{i=1}^{N} \widetilde{w}_{k+1|n}^{(i)} \frac{p\left(\mathbf{x}_{k+1}^{(i)} \middle| \mathbf{x}_k\right)}{p\left(\mathbf{x}_{k+1}^{(i)} \middle| \mathbf{y}_{0:k}\right)} \tag{58}$$

where, owing to (48), $p\left(\left.\mathbf{x}_{k+1}^{(i)}\right|\mathbf{y}_{0:k}\right)$ can be approximated by

$$p\left(\left.\mathbf{x}_{k+1}^{(i)}\right|\mathbf{y}_{0:k}\right) = \int p\left(\left.\mathbf{x}_{k+1}^{(i)}\right|\mathbf{x}_k\right) p\left(\left.\mathbf{x}_k\right|\mathbf{y}_{0:k}\right) d\mathbf{x}_k \simeq \sum_{j=1}^{N} \widetilde{w}_k^{(j)} p\left(\left.\mathbf{x}_{k+1}^{(i)}\right|\mathbf{x}_k^{(j)}\right) \tag{59}$$

An approximation $\widehat{P}\left(\left.d\mathbf{x}_k\right|\mathbf{y}_{0:n}\right)$ of $p\left(\left.\mathbf{x}_k\right|\mathbf{y}_{0:n}\right)$ is thus

$$\begin{aligned}
&\widehat{P}\left(\left.d\mathbf{x}_k\right|\mathbf{y}_{0:n}\right) \\
&= \left[\sum_{i=1}^{N} \widetilde{w}_k^{(i)} \delta_{\mathbf{x}_k^{(i)}}\left(d\mathbf{x}_k\right)\right] \sum_{j=1}^{N} \widetilde{w}_{k+1|n}^{(j)} \frac{p\left(\left.\mathbf{x}_{k+1}^{(j)}\right|\mathbf{x}_k\right)}{\left[\sum_{l=1}^{N} \widetilde{w}_k^{(l)} p\left(\left.\mathbf{x}_{k+1}^{(j)}\right|\mathbf{x}_k^{(l)}\right)\right]} \\
&= \sum_{i=1}^{N} \widetilde{w}_k^{(i)} \left[\sum_{j=1}^{N} \widetilde{w}_{k+1|n}^{(j)} \frac{p\left(\left.\mathbf{x}_{k+1}^{(j)}\right|\mathbf{x}_k^{(i)}\right)}{\left[\sum_{l=1}^{N} \widetilde{w}_k^{(l)} p\left(\left.\mathbf{x}_{k+1}^{(j)}\right|\mathbf{x}_k^{(l)}\right)\right]}\right] \delta_{\mathbf{x}_k^{(i)}}\left(d\mathbf{x}_k\right) \\
&\triangleq \sum_{i=1}^{N} \widetilde{w}_{k|n}^{(i)} \delta_{\mathbf{x}_k^{(i)}}\left(d\mathbf{x}_k\right)
\end{aligned} \tag{60}$$

The algorithm follows.

This algorithm requires storage of the marginal distributions $\widehat{P}\left(\left.d\mathbf{x}_k\right|\mathbf{y}_{0:k}\right)$ (weights and supports) for any $k = 0, ..., n$. The memory requirement is $O(nN)$. Its complexity is $O\left(nN^2\right)$, which is quite important as $N \gg 1$. However this complexity is a little lower than the one of the previous developed algorithms of (Kitagawa et al., 1996) and (Tanizaki et al., 1998) as it does not require any new simulation step.

D. Likelihood

In some applications, in particular for model choice (Kitagawa, 1987)(Kitagawa et al., 1996), we may wish to estimate the likelihood of the data

$$p\left(\mathbf{y}_{0:n}\right) = \int w_n^* \pi\left(\left.\mathbf{x}_{0:n}\right|\mathbf{y}_{0:n}\right) d\mathbf{x}_{0:n}$$

A simple estimate of the likelihood is thus given by

$$\widehat{p}\left(\mathbf{y}_{0:n}\right) = \frac{1}{N}\sum_{j=1}^{N} w_n^{(j)} \tag{61}$$

In practice, the introduction of resampling steps makes this approach impossible. We will use an alternative decomposition of the likelihood:

$$p(\mathbf{y}_{0:n}) = p(\mathbf{y}_0) \prod_{k=1}^{n} p(\mathbf{y}_k | \mathbf{y}_{0:k-1}) \tag{62}$$

where:

$$\begin{aligned} p(\mathbf{y}_k | \mathbf{y}_{0:k-1}) &= \int p(\mathbf{y}_k | \mathbf{x}_k) \, p(\mathbf{x}_k | \mathbf{y}_{0:k-1}) \, d\mathbf{x}_k &(63)\\ &= \int p(\mathbf{y}_k | \mathbf{x}_{k-1}) \, p(\mathbf{x}_{k-1} | \mathbf{y}_{0:k-1}) \, d\mathbf{x}_{k-1} &(64) \end{aligned}$$

Using (63), an estimate of this quantity is given by

$$\widehat{p}(\mathbf{y}_k | \mathbf{y}_{0:k-1}) = \sum_{i=1}^{N} p\left(\mathbf{y}_k | \widetilde{\mathbf{x}}_k^{(i)}\right) \widetilde{w}_{k-1}^{(i)} \tag{65}$$

where the samples $\left\{\widetilde{\mathbf{x}}_k^{(i)}; i = 1, \ldots, N\right\}$ are obtained using a one-step ahead prediction based on the approximation $\widehat{P}(d\mathbf{x}_{k-1} | \mathbf{y}_{0:k-1})$ of $p(\mathbf{x}_{k-1} | \mathbf{y}_{0:k-1})$. Using expression (64), it is possible to avoid a MC integration if we know analytically $p\left(\mathbf{y}_k | \mathbf{x}_{k-1}^{(i)}\right)$:

$$\widehat{p}(\mathbf{y}_k | \mathbf{y}_{0:k-1}) = \sum_{i=1}^{N} p\left(\mathbf{y}_k | \mathbf{x}_{k-1}^{(i)}\right) \widetilde{w}_{k-1}^{(i)} \tag{66}$$

VI. Simulations

In this section, we apply the methods developed previously to a linear Gaussian state space model and to a classical nonlinear model. We make for these two models $M = 100$ simulations of length $n = 500$ and we evaluate the empirical standard deviation for the filtering estimates $\mathbf{x}_{k|k} = \mathbb{E}[\mathbf{x}_k | \mathbf{y}_{0:k}]$ obtained by the MC methods:

$$\sqrt{VAR}(\mathbf{x}_{k|l}) = \frac{1}{n} \sum_{k=1}^{n} \left(\frac{1}{M} \sum_{j=1}^{M} \left(\mathbf{x}_{k|l}^{j} - \mathbf{x}_k^{j} \right)^2 \right)^{1/2}$$

where:

- $\mathbf{x}_k^j$ is the simulated state for the j^{th} simulation, $j = 1, ..., M$.

- $\mathbf{x}_{k|l}^{j} \triangleq \sum_{i=1}^{N} \widetilde{w}_{k|l}^{(i)} \mathbf{x}_{k}^{j,(i)}$ is the MC estimate of $\mathbb{E}[\mathbf{x}_k | \mathbf{y}_{0:l}]$ for the j^{th} test signal and $\mathbf{x}_{k}^{j,(i)}$ is the i^{th} simulated trajectory, $i = 1, ..., N$, associated with the signal j. (We denote $\widetilde{w}_{k|k}^{(i)} \triangleq \widetilde{w}_{k}^{(i)}$)

These calculations have been realised for $N = 100$, 250, 500, 1000, 2500 and 5000. The implemented filtering algorithms are the bootstrap filter, the SIS with the prior importance function and the SIS with the optimal or a suboptimal importance function. The fixed-interval smoothers associated with these SIS filters are then computed.

For the SIS-based algorithms, the SIR procedure has been used when $\widehat{N_{eff}} < N_{thres} = N/3$. We state the percentage of iterations where the SIR step is used for each importance function.

A. Linear Gaussian model

Let us consider the following model

$$x_k = x_{k-1} + v_k \tag{67}$$

$$y_k = x_k + w_k \tag{68}$$

where $x_0 \sim \mathcal{N}(0,1)$, v_k and w_k are white Gaussian noises mutually independent, $v_k \sim \mathcal{N}\left(0, \sigma_v^2\right)$ and $w_k \sim \mathcal{N}\left(0, \sigma_w^2\right)$ with $\sigma_v^2 = \sigma_w^2 = 1$. For this model, the optimal filter is the Kalman filter (Anderson et al., 1979).

1. Optimal importance function

The optimal importance function is

$$x_k | x_{k-1}, y_k \sim \mathcal{N}\left(m_k, \sigma_k^2\right) \tag{69}$$

where

$$\sigma_k^{-2} = \sigma_w^{-2} + \sigma_v^{-2} \tag{70}$$

$$m_k = \sigma_k^2 \left(\frac{x_{k-1}}{\sigma_v^2} + \frac{y_k}{\sigma_w^2} \right) \tag{71}$$

and the associated importance weight is equal to:

$$p\left(y_k|\,x_{k-1}\right) \propto \exp\left(-\frac{1}{2}\frac{\left(y_k - x_{k-1}\right)^2}{\left(\sigma_v^2 + \sigma_w^2\right)}\right) \tag{72}$$

2. Results

For the Kalman filter, we obtain $\sqrt{VAR}\left(\mathbf{x}_{k|k}\right) = 0.79$. For the different MC filters, the results are presented in Table 1 and Table 2.

With $N = 500$ trajectories, the estimates obtained using MC methods are similar to those obtained by Kalman. The SIS algorithms have similar performances to the bootstrap filter for a smaller computational cost. The most interesting algorithm is based on the optimal importance function which limits seriously the number of resampling steps.

B. Nonlinear series

We consider here the following nonlinear reference model (Gordon et al., 1993)(Kitagawa, 1987)(Tanizaki et al., 1998):

$$\begin{aligned} x_k &= f\left(x_{k-1}\right) + v_k & (73)\\ &= \frac{1}{2}x_{k-1} + 25\frac{x_{k-1}}{1+\left(x_{k-1}\right)^2} + 8\cos\left(1.2k\right) + v_k \\ y_k &= g\left(x_k\right) + w_k & (74)\\ &= \frac{\left(x_k\right)^2}{20} + w_k \end{aligned}$$

where $x_0 \sim \mathcal{N}(0,5)$, v_k and w_k are mutually independent white Gaussian noises, $v_k \sim \mathcal{N}\left(0,\sigma_v^2\right)$ and $w_k \sim \mathcal{N}\left(0,\sigma_w^2\right)$ with $\sigma_v^2 = 10$ and $\sigma_w^2 = 1$. In this case, it is not possible to evaluate analytically $p\left(y_k|\,x_{k-1}\right)$ or to sample simply from $p\left(x_k|\,x_{k-1}, y_k\right)$. We propose to apply the method described in 2. which consists of linearising locally the observation equation.

1. Importance function obtained by local linearisation

We get

$$
\begin{aligned}
y_k &\simeq g\left(f\left(x_{k-1}\right)\right) + \left.\frac{\partial g\left(x_k\right)}{\partial x_k}\right|_{x_k = f\left(x_{k-1}\right)} \left(x_k - f\left(x_{k-1}\right)\right) + w_k \\
&= \frac{f^2\left(x_{k-1}\right)}{20} + \frac{f\left(x_{k-1}\right)}{10}\left(x_k - f\left(x_{k-1}\right)\right) + w_k \\
&= -\frac{f^2\left(x_{k-1}\right)}{20} + \frac{f\left(x_{k-1}\right)}{10} x_k + w_k \qquad (75)
\end{aligned}
$$

Then we obtain the linearised importance function $\pi\left(x_k \mid x_{k-1}, y_k\right) = \mathcal{N}\left(x_k; m_k, \left(\sigma_k\right)^2\right)$ where

$$
\left(\sigma_k\right)^{-2} = \sigma_v^{-2} + \sigma_w^{-2} \frac{f^2\left(x_{k-1}\right)}{100} \qquad (76)
$$

and

$$
m_k = \left(\sigma_k\right)^2 \left[\sigma_v^{-2} f\left(x_{k-1}\right) + \sigma_w^{-2} \frac{f\left(x_{k-1}\right)}{10}\left(y_k + \frac{f^2\left(x_{k-1}\right)}{20}\right)\right] \qquad (77)
$$

2. Results

In this case, it is not possible to estimate the optimal filter. For the MC filters, the results are displayed in Table 3. The average percentages of SIR steps are presented in Table 4.

This model requires simulation of more samples than the preceding one. In fact, the variance of the dynamic noise is more important and more trajectories are necessary to explore the space. The most interesting algorithm is the SIS with a suboptimal importance function which greatly limits the number of resampling steps over the prior importance function while avoiding a MC integration step needed to evaluate the optimal importance function. This can be roughly explained by the fact that the observation noise is rather small so that y_k is highly informative and allows a limitation of the regions explored.

VII. Conclusion

We have presented an overview of sequential simulation-based methods for Bayesian filtering of general state-space models. We include, within the general framework of SIS, numerous approaches proposed independently in the literature over the last 30 years. Several original extensions have also been described, including the use of local linearisation techniques to yield more effective importance distributions. We have shown also how the use of Rao-Blackwellisation allows us to make the most of any analytic structure present in some important dynamic models and have described procedures for prediction, fixed-lag smoothing and likelihood evaluation.

These methods are efficient but still suffer from several drawbacks. The first is the depletion of samples which inevitably occurs in all of the methods described as time proceeds. Sample regeneration methods based upon MCMC steps are likely to improve the situation here (MacEachern et al. 1998). A second problem is that of simulating fixed hyperparameters such as the covariance matrices and noise variances which were assumed known in our examples. The methods described here do not allow for any regeneration of new values for these non-dynamic parameters, and hence we can expect a very rapid impoverishment of the sample set. Again, a combination of the present techniques with MCMC steps could be useful here, as could Rao-Blackwellisation methods ((Liu et al. 1998) give some insight into how this might be approached).

The technical challenges still posed by this problem, together with the wide range of important applications and the rapidly increasing computational power, should stimulate new and exciting developments in the field.

References

[1] Akashi H. and Kumamoto H.(1975) Construction of Discrete-time Nonlinear Filter by Monte Carlo Methods with Variance-reducing Techniques. *Systems and Control*, 19, 211-221 (in Japanese).

[2] Akashi H. and Kumamoto H.(1977) Random Sampling Approach to State Estimation

in Switching Environments. *Automatica*, 13, 429-434.

[3] Anderson B.D.O. and Moore J.B. (1979) *Optimal Filtering*, Englewood Cliffs.

[4] Berzuini C., Best N., Gilks W. and Larizza C. (1997) Dynamic Conditional Independence Models and Markov Chain Monte Carlo Methods. *Journal of the American Statistical Association*, 92, pp.1403-1412.

[5] Billio M. and Monfort A. (1998) Switching State-Space Models:Likelihood Function, Filtering and Smoothing. *Journal of Statistical Planning and Inference,* 68, pp.65-103.

[6] Carpenter J., Clifford P. and Fearnhead P. (1997) An Improved Particle Filter for Non-linear Problems. Technical report University of Oxford, Dept. of Statistics.

[7] Casella G. and Robert C.P. (1996) Rao-Blackwellisation of Sampling Schemes. *Biometrika*, 83, pp. 81-94.

[8] Chen R. and Liu J.S. (1996) Predictive Updating Methods with Application to Bayesian Classification. *Journal of the Royal Statistical Society B*, 58, 397-415.

[9] Clapp T.C. and Godsill S.J. (1999) Fixed-Lag Smoothing using Sequential Importance Sampling. forthcoming *Bayesian Statistics 6*, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), Oxford University Press.

[10] Doucet A. (1997) *Monte Carlo Methods for Bayesian Estimation of Hidden Markov Models. Application to Radiation Signals*. Ph.D. Thesis, University Paris-Sud Orsay (in French).

[11] Doucet A. (1998) On Sequential Simulation-Based Methods for Bayesian Filtering. Technical report University of Cambridge, Dept. of Engineering, CUED-F-ENG-TR310. Available on the MCMC preprint service at `http://www.stats.bris.ac.uk/MCMC/`.

[12] Geweke J. (1989) Bayesian Inference in Econometrics Models using Monte Carlo Integration. *Econometrica*, 57, 1317-1339.

[13] Godsill S.J and Rayner P.J.W. (1998) *Digital Audio Restoration - A Statistical Model-Based Approach*, Springer.

[14] Gordon N.J., Salmond D.J. and Smith A.F.M. (1993) Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. *IEE-Proceedings-F*, 140, 107-113.

[15] Gordon N.J. (1997) A Hybrid Bootstrap Filter for Target Tracking in Clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 33, 353-358.

[16] Handschin J.E. and Mayne D.Q. (1969) Monte Carlo Techniques to Estimate the Conditional Expectation in Multi-stage Non-linear Filtering. *International Journal of Control*, 9, 547-559.

[17] Handschin J.E. (1970) Monte Carlo Techniques for Prediction and Filtering of Non-Linear Stochastic Processes. *Automatica*, 6, 555-563.

[18] Higuchi T. (1997) Monte Carlo Filtering using the Genetic Algorithm Operators. *Journal of Statistical Computation and Simulation*, 59, 1-23.

[19] Jazwinski A.H. (1970) *Stochastic Processes and Filtering Theory*, Academic Press.

[20] Kitagawa G. (1987) Non-Gaussian State-Space Modeling of Nonstationary Time Series. *Journal of the American Statistical Association*, 82, 1032-1063.

[21] Kitagawa G. and Gersch G. (1996) *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics, 116, Springer.

[22] Kong A., Liu J.S. and Wong W.H. (1994) Sequential Imputations and Bayesian Missing Data Problems. *Journal of the American Statistical Association*, 89, 278-288.

[23] Liu J.S. and Chen R.(1995) Blind Deconvolution via Sequential Imputation. *Journal of the American Statistical Association*, 90, 567-576.

[24] Liu J.S. (1996) Metropolized Independent Sampling with Comparison to Rejection Sampling and Importance Sampling. *Statistics and Computing*, 6, 113-119.

[25] Liu J.S. and Chen R. (1998) Sequential Monte Carlo Methods for Dynamic Systems. *Journal of the American Statistical Association*, 93, 1032-1044.

[26] MacEachern S.N, Clyde M. and Liu J.S. (1998), Sequential Importance Sampling for Nonparametric Bayes Models: The Next Generation, forthcoming *Canadian Journal of Statistics*.

[27] Müller P. (1991) Monte Carlo Integration in General Dynamic Models. *Contemporary Mathematics*, 115, 145-163.

[28] Müller P. (1992) Posterior Integration in Dynamic Models. *Computing Science and Statistics*, 24, 318-324.

[29] Pitt M.K. and Shephard N. (1999) Filtering via Simulation:Auxiliary Particle Filters. forthcoming *Journal of the American Statistical Association.*

[30] Ripley B.D., *Stochastic Simulation*, Wiley, New York, 1987.

[31] Rubin D.B. (1988) Using the SIR Algorithm to Simulate Posterior Distributions. in *Bayesian Statistics 3* (Eds J.M. Bernardo, M.H. DeGroot, D.V. Lindley et A.F.M. Smith), Oxford University Press, 395-402.

[32] Smith A.F.M. and Gelfand A.E. (1992) Bayesian Statistics without Tears: a Sampling-Resampling Perspective. *The American Statistician,* 46, 84-88.

[33] Stewart L. and McCarty P. (1992) The Use of Bayesian Belief Networks to Fuse Continuous and Discrete Information for Target Recognition, Tracking and Situation Assessment. *Proceeding Conference SPIE*, 1699, 177-185.

[34] Svetnik V.B. (1986) Applying the Monte Carlo Method for Optimum Estimation in Systems with Random Disturbances. *Automation and Remote Control*, 47, 818-825.

[35] Tanizaki H. (1993) *Nonlinear Filters: Estimation and Applications*, Lecture Notes in Economics and Mathematical Systems, 400, Springer, Berlin.

[36] Tanizaki H. and Mariano R.S. (1994) Prediction, Filtering and Smoothing in Non-linear and Non-normal Cases using Monte Carlo Integration. *Journal of Applied Econometrics*, 9, 163-179.

[37] Tanizaki H. and Mariano R.S. (1998) Nonlinear and Non-Gaussian State-Space Modeling with Monte-Carlo Simulations. *Journal of Econometrics,* 83, 263-290.

[38] Tugnait J.K. (1982) Detection and Estimation for Abruptly Changing Systems. *Automatica*, 18, 607-615.

[39] West M. (1993) Mixtures Models, Monte Carlo, Bayesian Updating and Dynamic Models. *Computer Science and Statistics*, 24, 325-333.

[40] West M. and Harrison J.F. (1997) *Bayesian Forecasting and Dynamic Models*, Springer Verlag Series in Statistics, 2nd edition.

[41] Zaritskii V.S., Svetnik V.B. and Shimelevich L.I. (1975) Monte Carlo Technique in Problems of Optimal Data Processing. *Automation and Remote Control*, 12, 95-103.

VIII. Tables

$\sqrt{VAR}\left(\mathbf{x}_{k\|k}\right)$	bootstrap	prior dist.	optimal dist.
$N = 100$	0.80	0.86	0.83
$N = 250$	0.81	0.81	0.80
$N = 500$	0.79	0.80	0.79
$N = 1000$	0.79	0.79	0.79
$N = 2500$	0.79	0.79	0.79
$N = 5000$	0.79	0.79	0.79

Table 1: MC filters: linear Gaussian model

Percentage SIR	prior dist.	optimal dist.
$N = 100$	40	16
$N = 250$	23	10
$N = 500$	20	8
$N = 1000$	15	6
$N = 2500$	13	5
$N = 5000$	11	4

Table 2: Percentage of SIR steps: linear Gaussian model

$\sqrt{VAR}\left(\mathbf{x}_{k\|k}\right)$	bootstrap	prior dist.	linearised dist.
$N = 100$	5.67	6.01	5.54
$N = 250$	5.32	5.65	5.46
$N = 500$	5.27	5.59	5.23
$N = 1000$	5.11	5.36	5.05
$N = 2500$	5.09	5.14	5.02
$N = 5000$	5.04	5.07	5.01

Table 3: MC filters: nonlinear time series

Percentage SIR	prior dist.	linearised dist.
$N = 100$	22.4	8.9
$N = 250$	19.6	7.5
$N = 500$	17.7	6.5
$N = 1000$	15.6	5.9
$N = 2500$	13.9	5.2
$N = 5000$	12.3	5.3

Table 4: Percentage of SIR steps: nonlinear time series