A Backward-Simulation Based Rao-Blackwellized Particle Smoother for Conditionally Linear Gaussian Models

Simo Särkkä * **Pete Bunch** ** **Simon J. Godsill** ***

* *Aalto University, P.O. Box 12200. FI-00076 AALTO, Finland. (Tel: +358 50 512 4393; e-mail: simo.sarkka@aalto.fi)*
** *University of Cambridge, Department of Engineering, Trumpington Street, Cambridge CB2 1PZ, UK (e-mail: pb404@cam.ac.uk)*
*** *University of Cambridge, Department of Engineering, Trumpington Street, Cambridge CB2 1PZ, UK (e-mail: sjg@cam.ac.uk)*

Abstract: In this article, we develop a new Rao-Blackwellized Monte Carlo smoothing algorithm for conditionally linear Gaussian models. The algorithm is based on the forward-filtering backward-simulation Monte Carlo smoother concept and performs the backward simulation directly in the marginal space of the non-Gaussian state component while treating the linear part analytically. Unlike the previously proposed backward-simulation based Rao-Blackwellized smoothing approaches, it does not require sampling of the Gaussian state component and is also able to overcome certain normalization problems of two-filter smoother based approaches. The performance of the algorithm is illustrated in a simulated application.

Keywords: estimation algorithms; optimal estimation; nonlinear systems; Monte Carlo method; stochastic systems.

1. INTRODUCTION

In this paper we derive a novel Rao-Blackwellized particle smoother for conditionally linear Gaussian models [see Doucet et al., 2000] of the form

$$\begin{aligned} u_k &\sim p(u_k \mid u_{k-1}) \\ z_k &= A(u_{k-1})\, z_{k-1} + q_{k-1} \\ y_k &= H(u_k)\, z_k + r_k, \end{aligned} \tag{1}$$

where the Gaussian process noise $q_{k-1} \sim \mathrm{N}(0, Q(u_{k-1}))$ and observation noise $r_k \sim \mathrm{N}(0, R(u_k))$ are assumed to be time-white. The state consists of two parts $x_k = (u_k, z_k)$, where $z_k \in \mathbb{R}^{n_z}$ is conditionally Gaussian given the non-Gaussian part $u_k \in \mathbb{R}^{n_u}$. Here we formally treat the variable u_k as having a probability density with respect to the Lebesgue measure, but all the results apply or can be easily generalized to random variables without such densities.

The key feature of the novel smoother is that it does not require sampling of the linear portion of the state z_k. This feature is required for the smoother to be truly Rao-Blackwellized (marginalized) as opposed to, for example, smoothers presented by Fong et al. [2002] and Lindsten and Schön [2011], which also sample the linear part. The proposed smoother is based on the forward-filtering backward-sampling approach of Godsill et al. [2004].

The structure of the paper is as follows: In Section 2 we briefly review the background theory of particle filtering, particle smoothing, and Rao-Blackwellization. In Section 3 we derive the novel Rao-Blackwellized particle smoother algorithm and finally in Section 4 we demonstrate the performance of the proposed algorithm in a simulated application.

2. BACKGROUND

2.1 Particle filters and smoothers

The theory of optimal Bayesian filtering and smoothing [Ho and Lee, 1964, Lee, 1964] is considered with state estimation in probabilistic state space models of the form

$$\begin{aligned} x_k &\sim p(x_k \mid x_{k-1}) \\ y_k &\sim p(y_k \mid x_k), \end{aligned} \tag{2}$$

where $k = 1, \ldots, n$, $x_k \in \mathbb{R}^{d_x}$ is the state of the system, $y_k \in \mathbb{R}^{d_y}$ is the observation at time step k, $p(x_k \mid x_{k-1})$ is the Markovian dynamic model probability density and $p(y_k \mid x_k)$ is the measurement model probability density.

As the Bayesian filtering equations do not in general admit closed form solutions, approximations are needed. A general class of these is the particle filter [Gordon et al., 1993, Kitagawa, 1996, Doucet et al., 2000, Ristic et al., 2004, Cappé et al., 2007], which generates a Monte Carlo approximation to the filtering distribution using the sequential importance resampling (SIR) algorithm. Formally this means the following approximation to the probability density, which consists of a weighted sum of Dirac delta functions:

$$p(x_k \,|\, y_{1:k}) \approx \sum_i w_k^{(i)} \delta(x_k - x_k^{(i)}). \tag{3}$$

Analogously to particle filters, particle smoothers can be used to form approximate solutions to the Bayesian smoothing equations, which also do not admit closed form

solutions in general. The particle smoother of Kitagawa [1996] is based on direct use of SIR for smoothing: If we simply store the entire sample histories and resample these instead of single-step samples, we obtain a smoothing solution alongside the filtering solution.

The problem with this approach is that its estimate of the smoothing distribution tends to be quite degenerate [Kitagawa, 1996]. For this reason, various alternative particle smoothers have been proposed. One possible approach is to approximate the two-filter formula using particle methods [Kitagawa, 1996, Briers et al., 2010, Fearnhead et al., 2010]. The disadvantage of such two-filter type smoothers is that the construction of the backward filter can be quite troublesome. Another way is to compute new weights for the particle filter results using an additional backward sweep, as is done in particle smoothers of Hürzeler and Kunsch [1998] and Doucet et al. [2000]. Because these approaches only approximate the marginal smoothing distributions without generating explicit trajectories from the joint smoothing distribution, their usage in Rao-Blackwellized particle smoothing is not straightforward.

In this article we shall use the backward-simulation based particle smoothing approach proposed by Godsill et al. [2004]. We first run a particle filter over the measurement sequence and store all the intermediate results. State trajectories are then generated by first sampling $\tilde{x}_n$ from the filtering distribution, then $\tilde{x}_{n-1}$ conditional on the sampled $\tilde{x}_n$, and so on. The approach works because of the following simple Markovian property of the model:

$$p(x_k \,|\, x_{k+1:n}, y_{1:n}) \propto p(x_{k+1} \,|\, x_k)\, p(x_k \,|\, y_{1:k}). \qquad (4)$$

Furthermore, because $p(x_k \,|\, y_{1:k})$ is the filtering distribution at step k, we can draw from the above distribution simply by drawing a filter sample $x_k^{(i)}$ with probability $p(\tilde{x}_{k+1} \,|\, x_k^{(i)})\, w_k^{(i)}$. Thus the algorithm is:

Algorithm 2.1. (Backward-simulation smoother). Given the weighted set of particles $\{w_k^{(i)}, x_k^{(i)} \mid i = 1, \ldots, N,\ k = 1, \ldots, n\}$ representing the filtering distributions:

- Choose $\tilde{x}_n = x_n^{(i)}$ with probability $w_n^{(i)}$.
- For $k = n-1, \ldots, 0$:
  (1) Compute new weights by
  $$w_{k|k+1}^{(i)} \propto w_k^{(i)} p(\tilde{x}_{k+1} \,|\, x_k^{(i)}) \qquad (5)$$
  (2) Choose $\tilde{x}_k = x_k^{(i)}$ with probability $w_{k|k+1}^{(i)}$

Given S iterations of the above procedure resulting in samples $\tilde{x}_{1:n}^{(j)}$ for $j = 1, \ldots, S$ the smoothing distribution can now be approximated as

$$p(x_{1:n} \,|\, y_{1:n}) \approx \frac{1}{S} \sum_j \delta(x_{1:n} - \tilde{x}_{1:n}^{(j)}). \qquad (6)$$

2.2 Rao-Blackwellized particle filters and smoothers

Although the generic particle filters and smoothers can be used in almost any kind of models, the required number of samples for a sufficient accuracy can be high. The efficiency of sampling can be improved by Rao-Blackwellization [Doucet et al., 2000], where part of the state is marginalized out in closed form, and only the remaining part is sampled. Because the sampled space has a lower dimension, fewer particles are required. Such closed form marginalization is possible, for example, in conditionally linear Gaussian models in Equations (1).

The Rao-Blackwellized particle filter (RBPF) [Doucet et al., 2000, Chen and Liu, 2000] is the algorithm which results if we marginalize the linear state z_k from the conditionally linear Gaussian model (1), and use SIR for the non-Gaussian part u_k only. At each step the approximation to the filtering density is a mixture of Gaussians:

$$p(u_k, z_k \,|\, y_{1:k}) = \sum_i w_k^{(i)} \mathrm{N}(z_k \,|\, m_k^{(i)}, P_k^{(i)})\, \delta(u_k - u_k^{(i)}),$$

where $u_k^{(i)}$ are the latent variable samples and $m_k^{(i)}$, $P_k^{(i)}$ are the mean and covariance of the Kalman filter conditioned on the corresponding history of latent variables.

Analogously to the non-Rao-Blackwellized case, we can obtain a Kitagawa [1996] type of approximation to the smoothing solution along with the filtering by storing and resampling the histories $u_{1:k}^{(i)}$ and the corresponding Kalman filter means and covariances as well. The marginal smoothing solution for $u_{1:k}^{(i)}$ is then the following:

$$p(u_{1:k} \,|\, y_{1:k}) \approx \sum_i w_k^{(i)} \delta(u_{1:k} - u_{1:k}^{(i)}). \qquad (7)$$

The approximation to the linear part of the smoothing distribution can then be obtained by running Rauch-Tung-Striebel (RTS) smoothers for each of the histories to yield smoothed means and covariances $m_{1:n}^{s,(i)}$ and $P_{1:n}^{s,(i)}$. The approximation to the marginal smoothing density at step k is then

$$p(u_k, z_k \,|\, y_{1:n}) = \sum_i w_k^{(i)} \mathrm{N}(z_k \,|\, m_k^{s,(i)}, P_k^{s,(i)})\, \delta(u_k - \tilde{u}_k^{(i)}),$$

where $\tilde{u}_k^{(i)}$ is the ith component of the history $u_{1:k}^{(i)}$.

A class of simple approximate Rao-Blackwellized particle smoothers (RBPSs) can be obtained by using *Kim's approximation* [Kim, 1994, Barber, 2006]:

$$\begin{aligned} &p(u_k \,|\, u_{k+1:n}, y_{1:n}) \\ &= \int p(u_k \,|\, u_{k+1}, z_{k+1}, y_{1:k})\, p(z_{k+1} \,|\, u_{k+1}, y_{1:n})\, dz_{k+1} \qquad (8) \\ &\approx p(u_k \,|\, u_{k+1}, y_{1:k}). \end{aligned}$$

In other words, we smooth the non-Gaussian variables u_k by ignoring the effect of Gaussian variables z_k completely. The backward-simulation smoother based on Kim's approximation amounts to replacing the transition density $p(x_{k+1} \,|\, x_k)$ in the Algorithm 2.1 with $p(u_{k+1} \,|\, u_k)$. Given a trajectory of the non-Gaussian variable, the linear Gaussian part may be estimated with Kalman filter and RTS smoother.

Another possible approach for the RBPS is to use the two-filter formula, and form the smoothing solution by combining two RBPFs, one running forward and one backwards in time [Briers et al., 2010, Fearnhead et al., 2010].

The RBPSs introduced by Fong et al. [2002] and Lindsten and Schön [2011] are based on simulating backward trajectories from the joint distribution (z_k, u_k). However, these smoothers are not really Rao-Blackwellized backward-

simulation smoothers, because they require sampling of the linear part of the state as well.

A related, but slightly different approach to inference in conditionally linear Gaussian models uses mixture-of-Gaussians approximations, presented, for example, in Kim [1994], Bar-Shalom et al. [2001], Barber [2006].

3. NOVEL RAO-BLACKWELLIZED BACKWARD-SIMULATION PARTICLE SMOOTHER

3.1 Idea of Marginal Simulation

In the Rao-Blackwellized case, we cannot do backward simulation just by drawing samples $u_k^{(i)}$ one step at a time from the filter results (as in Algorithm 2.1), because the linear part of the state (and thus the likelihood of the measurements) depends on the entire history $u_{1:k}^{(i)}$. Thus we need to sample whole trajectories from the sample histories produced by RB filter/smoother. Furthermore, the marginalized state u_k does not have a simple Markov property such as (4) and thus we need to condition on the whole future instead of just on the next time step. Thus, we must calculate weights for each filtering trajectory, $u_{1:k}^{(i)}$, when it is paired with the sampled (fixed) future state sequence, $\tilde{u}_{k+1:n}$.

In order to do backward sampling for $u_{1:n}$, we will need an efficient way to evaluate the following distribution for different values of $u_{0:k}$:

$$p(u_{0:k} \,|\, u_{k+1:n}, y_{1:n}) = \frac{p(u_{0:n} \,|\, y_{1:n})}{p(u_{k+1:n} \,|\, y_{1:n})}. \tag{9}$$

Because we are sampling a single trajectory, we can consider the denominator in (9) as fixed. Thus we are interested in evaluating

$$p(u_{0:n} \,|\, y_{1:n}) \propto p(y_{1:n} \,|\, u_{0:n})\, p(u_{0:n}). \tag{10}$$

The prior $p(u_{0:n}) = p(u_0) \prod_{j=1}^{n} p(u_j \,|\, u_{j-1})$ is easy to evaluate; the problematic term is the marginal likelihood $p(y_{1:n} \,|\, u_{0:n})$. It could be calculated in a brute force way, by running a Kalman filter through the data. However, at each sampling time k we then would need to run the Kalman filters in each particle i over the entire sampled future $\tilde{u}_{k+1:n}$. Thus we need a more efficient approach.

3.2 Recursion Equations

We can split up the dependency between the past and future in the marginal likelihood by augmenting with z_k, which gives:

$$\begin{aligned} &p(y_{1:n}, z_k \,|\, u_{0:n}) \\ &= p(y_{k+1:n} \,|\, z_k, u_{0:n}, y_{1:k})\, p(y_{1:k}, z_k \,|\, u_{0:n}) \\ &= p(y_{k+1:n} \,|\, z_k, u_{k:n})\, p(z_k \,|\, y_{1:k}, u_{0:n})\, p(y_{1:k} \,|\, u_{0:n}) \\ &= p(y_{k+1:n} \,|\, z_k, u_{k:n})\, p(z_k \,|\, y_{1:k}, u_{0:k})\, p(y_{1:k} \,|\, u_{0:k}). \end{aligned} \tag{11}$$

The last two terms in Equation (11) are given by:

$$\begin{aligned} p(z_k \,|\, y_{1:k}, u_{0:k}) &= \mathrm{N}(z_k \,|\, m_k, P_k) \\ p(y_{1:k} \,|\, u_{0:k}) &= \prod_{j=1}^{k} \mathrm{N}(y_j \,|\, H(u_j)\, m_j^-, S_j), \end{aligned} \tag{12}$$

where the means and covariance can be computed with the Kalman filter recursions

$$\begin{aligned} m_k^- &= A(u_{k-1})\, m_{k-1} \\ P_k^- &= A(u_{k-1})\, P_{k-1} A^T(u_{k-1}) + Q(u_{k-1}) \\ S_k &= H(u_k)\, P_k^-\, H^T(u_k) + R(u_k) \\ K_k &= P_k^-\, H^T(u_k)\, S_k^{-1} \\ m_k &= m_k^- + K_k\, [y_k - H(u_k)\, m_k^-] \\ P_k &= P_k^- - K_k\, S_k\, K_k^T. \end{aligned} \tag{13}$$

The first term in Equation (11) can be computed with a backward Kalman filter. First assume that there exists a mean and covariance m_{k+1}^b, P_{k+1}^b and a normalization constant Z_{k+1} such that

$$p(y_{k+1:n} | z_{k+1}, u_{k+1:n}) = Z_{k+1} \mathrm{N}(z_{k+1} | m_{k+1}^b, P_{k+1}^b). \tag{14}$$

By the two-filter smoother equations we then get

$$\begin{aligned} &p(y_{k+1:n} \,|\, z_k, u_{k:n}) \\ &= \int p(y_{k+1:n} \,|\, z_{k+1}, u_{k+1:n})\, p(z_{k+1} \,|\, z_k, u_k)\, dz_{k+1} \\ &= Z_{k+1} |\det A(u_k)|^{-1} \mathrm{N}(z_k \,|\, m_k^{-b}, P_k^{-b}), \end{aligned} \tag{15}$$

where

$$\begin{aligned} m_k^{-b} &= A^{-1}(u_k)\, m_{k+1}^b \\ P_k^{-b} &= A^{-1}(u_k)\, (Q(u_k) + P_{k+1}^b)\, A^{-T}(u_k), \end{aligned} \tag{16}$$

and

$$\begin{aligned} &p(y_{k:n} \,|\, z_k, u_{k:n}) \\ &= p(y_k \,|\, z_k, u_{k:n})\, p(y_{k+1:n} \,|\, z_k, u_{k:n}) \\ &= Z_{k+1} |\det A(u_k)|^{-1}\, \mathrm{N}(y_k \,|\, \mu_k^b, S_k^b)\, \mathrm{N}(z_k \,|\, m_k^b, P_k^b) \\ &= Z_k\, \mathrm{N}(z_k \,|\, m_k^b, P_k^b), \end{aligned} \tag{17}$$

where

$$\begin{aligned} \mu_k^b &= H(u_k)\, m_k^{-b} \\ S_k^b &= H(u_k) P_k^{-b}\, H^T(u_k) + R(u_k) \\ K_k^b &= P_k^{-b}\, H^T(u_k)\, (S_k^b)^{-1} \\ m_k^b &= m_k^{-b} + K_k^b\, [y_k - \mu_k^b] \\ P_k^b &= P_k^{-b} - K_k^b\, S_k^b\, (K_k^b)^T \\ Z_k &= Z_{k+1} \,|\det A(u_k)|^{-1}\, \mathrm{N}(y_k | \mu_k^b, S_k^b). \end{aligned} \tag{18}$$

Substituting the results into Equation (11) and integrating over z_k then gives

$$\begin{aligned} &p(y_{1:n} \,|\, u_{0:n}) \\ &= Z_{k+1} \,|\det A(u_k)|^{-1} \mathrm{N}(m_k \,|\, m_k^{-b}, P_k + P_k^{-b}) \\ &\quad \times \prod_{j=1}^{k} \mathrm{N}(y_j \,|\, H(u_j)\, m_j^-, S_j). \end{aligned} \tag{19}$$

By combining with the prior of $u_{0:n}$ and ignoring the terms that only depend on $u_{k+1:n}$, we get

$$\begin{aligned} &p(u_{0:k} \,|\, u_{k+1:n}, y_{1:n}) \\ &\propto |\det A(u_k)|^{-1} \mathrm{N}(m_k \,|\, m_k^{-b}, P_k + P_k^{-b})\, p(u_{k+1} \,|\, u_k) \\ &\quad \times p(u_0) \prod_{j=1}^{k} p(u_j \,|\, u_{j-1})\, \mathrm{N}(y_j \,|\, H(u_j)\, m_j^-, S_j). \end{aligned} \tag{20}$$

The term on the second line is just the distribution approximated by the RBPF, (Eq. (7)). Substituting the particle approximation then gives

$$
\begin{aligned}
&p(u_{0:k} \,|\, u_{k+1:n}, y_{1:n}) \\
&\propto |\det A(u_k)|^{-1} \mathrm{N}(m_k \,|\, m_k^{-b}, P_k + P_k^{-b})\, p(u_{k+1} \,|\, u_k) \\
&\quad \times \left[\sum_{i=1}^{N} w_k^{(i)}\, \delta(u_{0:k} - u_{0:k}^{(i)}) \right].
\end{aligned} \tag{21}
$$

3.3 Initialization

The recursion in the previous section is based on the assumption that the normalization constant in the Equation (14) exists and is finite, making the "inversion" of the Gaussian density possible. Unfortunately, at the last step, when $k+1 = n$ it is often the case that such a normalization constant does not exist. This issue is related to the well-known normalization problem of two-filter smoothers. In the case of linear smoothing this problem can be overcome by using an information filter and defining the distribution at the last step as being formally singular. Unfortunately, although we only need to use a linear smoother, this approach does not solve the problem of undefined normalization constant in our case.

Fortunately, if the system is observable, the normalization constant is guaranteed to become finite at some point. One simple, approximate way to initialize the filter is to start by using the Kim's approximation based smoother until the normalization constant becomes finite (say s steps). An alternative, exact method is to calculate the first term of Equation (11) directly, using the observation density with an augmented vector of all the observations from $k+1$ to n. The form for this augmented observation density is given by Kitagawa [1994]. Thus, it is not necessary to invert the Gaussian to calculate the sampling weights until the normalization constant becomes finite.

3.4 Practical Implementation

The final algorithm is the following:

Algorithm 3.1. (Rao-Blackwellized smoother). Given the weighted set of particles $\{w_k^{(i)}, u_{1:k}^{(i)} \mid i = 1, \ldots, N,\ k = 1, \ldots, n\}$ representing the filtering results and their histories at different times:

- Choose $\tilde{u}_n = u_n^{(i)}$ with probability $w_n^{(i)}$.
- For $k = n-1, \ldots, 0$:
  (1) Compute new weights by
  $$
  \begin{aligned}
  w_{k|k+1}^{(i)} &\propto w_k^{(i)}\, p(\tilde{u}_{k+1} \,|\, u_k^{(i)})\, |\det A(u_k^{(i)})|^{-1} \\
  &\quad \times \mathrm{N}(m_k^{(i)} \,|\, m_k^{-b,(i)}, P_k^{(i)} + P_k^{-b,(i)}),
  \end{aligned} \tag{22}
  $$
  where $m_k^{(i)}$ and $P_k^{(i)}$ are the forward Kalman filter mean and covariance for the trajectory $u_{1:k}^{(i)}$, and $m_k^{-b,(i)}$ and $P_k^{-b,(i)}$ are the backward filter results backward predicted to step k using $u_k^{(i)}$.
  (2) Choose $\tilde{u}_k = u_k^{(i)}$ with probability $w_{k|k+1}^{(i)}$.
- The smoothing solution to the linear part can be computed by running a Kalman filter and RTS smoother over the generated trajectory $\tilde{u}_{1:n}$. In practice, we can also compute the solution during the backward simulation by combining the computed forward and backward filtering solutions using the two-filter smoother formulas.

Given S iterations of the above procedure resulting in samples $\tilde{u}_{1:n}^{(j)}$ for $j = 1, \ldots, S$ as well as the corresponding smoother means and covariances $m_{1:n}^{s,(j)}$, $P_{1:n}^{s,(j)}$, the smoothing distribution can be approximated as

$$
p(u_k, z_k \,|\, y_{1:n}) = \frac{1}{S} \sum_j \mathrm{N}(z_k \,|\, m_k^{s,(j)}, P_k^{s,(j)})\, \delta(u_k - \tilde{u}_k^{(j)}).
$$

4. NUMERICAL RESULTS

For testing the practical performance of the proposed algorithm, we used a model in which the linear Gaussian part of the state z_k is 2-dimensional and the non-Gaussian part is an indicator variable taking discrete values $c_k \in \{1, 2\}$. The model is the following:

- If the latent variable $c_{k-1} = 1$, then the dynamic model for the step $t_{k-1} \to t_k$ is the Wiener velocity model [cf. Bar-Shalom et al., 2001] $\ddot{z} = w(t)$, where $w(t)$ is a white noise process with spectral density $q_c = 0.01$
- If the latent variable $c_{k-1} = 2$ then the dynamic model is a damped harmonic force model $\ddot{z} = -z/10 - \dot{z}/10 + w(t)$.
- If $c_k = 1$ then we measure z with variance 1, otherwise with variance 5^2.

The transition probabilities for c_k were $\Pi_{1\,|\,1} = 0.8$, $\Pi_{1\,|\,2} = 0.2$, $\Pi_{2\,|\,1} = 0.2$, and $\Pi_{2\,|\,2} = 0.8$. The initial distributions were $P(c_0 = 1) = 0.9$, $P(c_0 = 2) = 0.1$ and $z_0 \sim \mathrm{N}((0\ 1)^T, I)$. The sampling period was $\Delta t = 0.1$ and the continuous time models were discretized using the standard procedure [see Bar-Shalom et al., 2001]. The "optimal importance distribution" [Doucet et al., 2000] was used as a proposal for the filter.

The following methods were tested:

- *RBPF*: Rao-Blackwellized particle filter.
- *KiS*: Rao-Blackwellized Kitagawa [1996] smoother.
- *KGS*: Kim's approximation based Rao-Blackwellized Godsill et al. [2004] smoother.
- *FGS*: Fong et al. [2002] smoother.
- *RBG*: The proposed Rao-Blackwellized Godsill et al. [2004] smoother in Algorithm 3.1.

Performance was assessed by comparison of the root mean square error in the linear part of the state (*RMSE*). In addition, at each time step a maximum a posteriori estimate for the value of the indicator variable was made by selecting the option with the greater weight. The proportion of errors in these estimates was then calculated (*ErrRate*). Furthermore, by averaging the total weight assigned to the unselected indicator values, we can calculate the error rate predicted by each algorithm (*PredRate*). A comparison of the predicted and observed error rates gives an additional measure of how accurately the algorithms characterize the posterior distribution.

The results from 1000 independent simulation runs using 100 particles in each of the methods are shown in Table 1. The RMSE results of the smoother methods are almost equal, except KiS gives a slightly higher RMSE on average than the other smoothers. The error rate of KiS is higher than of the other smoothers although the predicted error rate is significantly lower than of the others. The RMSEs,

error rates and predicted errors of the smoothers as a function of time are shown in Figures 1–3. It can be seen that RMSE and error rate of KiS increases (relative to other methods), and the predicted error rate decreases toward the beginning of the data. These effects are likely to be due to the degeneracy of KiS. The error rate and predicted error rate of KGS are slightly higher than those of FGS and RBG, whereas the results of the latter seem to be practically identical.

Method	RMSE	ErrRate	PredRate
RBPF	0.49	0.16	0.16
KiS	0.27	0.16	0.09
KGS	0.26	0.14	0.13
FGS	0.26	0.13	0.12
RBG	**0.26**	**0.13**	**0.12**

Table 1. Simulation results with 100 particles averaged over 1000 runs.

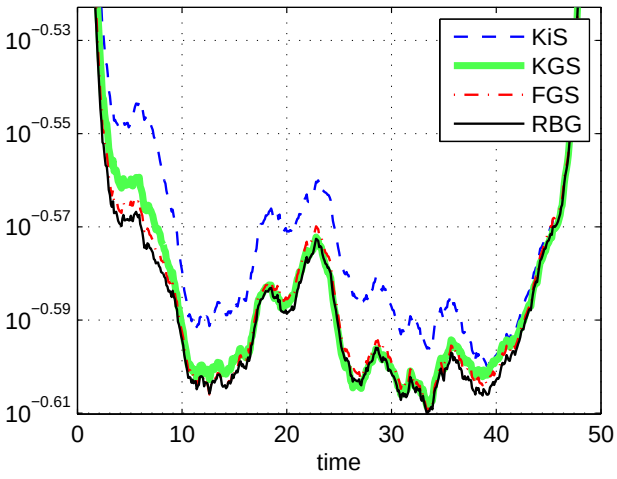

Fig. 1. RMSEs with 100 particles averaged over 1000 runs.

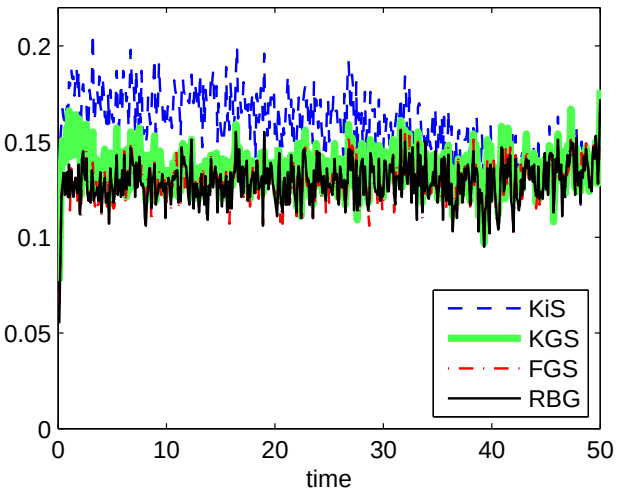

Fig. 2. Error rates with 100 particles averaged over 1000 runs.

To test the effect of using a small number of particles, we also ran 1000 simulation runs with only 10 particles. The results are shown in Table 2 and they seem to be consistent with the 100 particle results. Now the RMSE of

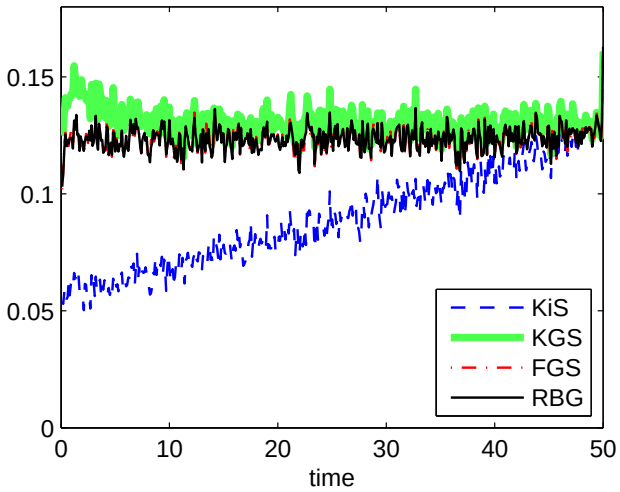

Fig. 3. Predicted error rates with 100 particles averaged over 1000 runs.

Method	RMSE	ErrRate	PredRate
RBPF	0.52	0.18	0.14
KiS	0.30	0.18	0.02
KGS	0.29	0.16	0.11
FGS	0.30	0.15	0.10
RBG	**0.29**	**0.15**	**0.10**

Table 2. Simulation results with 10 particles averaged over 1000 runs.

FGS is slightly higher on average than that of RBG. This is to be expected, because the linear state in FGS is now represented by 10 samples whereas RBG maintains the whole Gaussian distribution. The higher RMSE of FGS can also be clearly seen in the Figure 4.

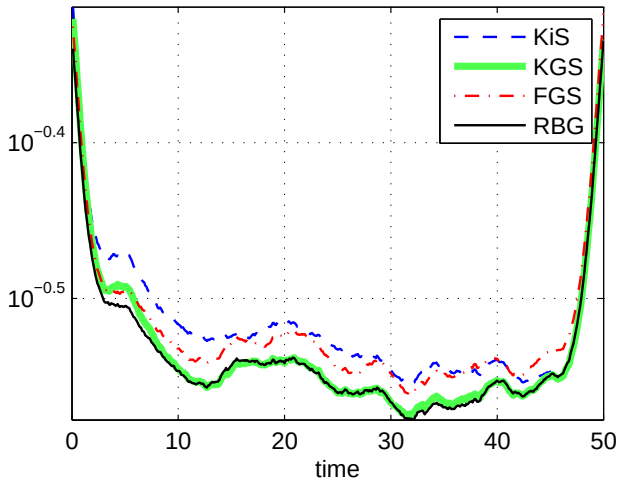

Fig. 4. RMSEs with 10 particles averaged over 1000 runs.

5. CONCLUSION AND DISCUSSION

In this article we have derived a novel Rao-Blackwellized particle smoother, which is based on the backward simulation approach of Godsill et al. [2004]. Unlike the schemes of Fong et al. [2002] and Lindsten and Schön [2011], the proposed smoother does not require sampling of the linear

part of the state. The resulting algorithm is similar to the pure particle smoother of Godsill et al. [2004], but now the computation of weights is a function of forward and backward Kalman filter estimates, conditioned on the sampled trajectory of the non-Gaussian part of the state.

The performance of the smoother was illustrated in a simulated application, which shows a slight improvement by the RBG over the other schemes. However, we expect to see a more significant difference when more complex models are used. In particular, the linear state component in the test presented is only two-dimensional, meaning that the distribution is well characterized by few samples. Furthermore, by using the optimal importance distribution, samples are in fact being drawn directly from the conditional posterior distribution, so very few are "wasted" in areas of low probability. If the linear state dimensionality were increased or no optimal importance distribution were available, more samples would be required for an accurate estimate. Thus we expect the RMSE performance of the FGS scheme to deteriorate faster than that of RBG. Such studies are currently the subject of further investigation.

ACKNOWLEDGEMENTS

Partially supported by the Academy of Finland. The work was done while the first author was a visiting scholar at the Signal Processing and Communications Laboratory, Department of Engineering, University of Cambridge. The computations were done using Aalto University's Triton cluster.

REFERENCES

Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. Wiley Interscience, 2001.

D. Barber. Expectation correction for smoothed inference in switching linear dynamical systems. *Journal of Machine Learning Research*, 7:2515–2540, 2006.

M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state-space models. *Ann Inst Stat Math*, 62: 61–89, 2010.

O. Cappé, S. J. Godsill, and E. Moulines. An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE*, 95(5):899–924, 2007.

R. Chen and J. S. Liu. Mixture Kalman filters. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62(3):493–508, 2000.

A. Doucet, S. J. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

P. Fearnhead, D. Wyncoll, and J. Tawn. A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2):447–464, 2010.

W. Fong, S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing with application to audio signal enhancement. *IEEE Transactions on Signal Processing*, 50(2):438–449, 2002.

S. J. Godsill, A. Doucet, and M. West. Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, 99(465):156–168, 2004.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F*, 140(2):107–113, 1993.

Y. C. Ho and R. C. K. Lee. A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, 9:333–339, 1964.

M. Hürzeler and H. R. Kunsch. Monte Carlo approximations for general state-space models. *Journal of Computational and Graphical Statistics*, 7(2):175–193, 1998.

C-J. Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60:1–22, 1994.

G. Kitagawa. The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother. *Annals of the Institute of Statistical Mathematics*, 46:605–623, 1994.

G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5:1–25, 1996.

R. C. K. Lee. *Optimal Estimation, Identification and Control*. M.I.T. Press, 1964.

F. Lindsten and T. B. Schön. Rao-Blackwellized particle smoothers for mixed linear/nonlinear state-space models. Technical Report LiTH-ISY-R-2018, Linköpings Universitet, 2011.

B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter*. Artech House, 2004.