An explicit variance reduction expression for the Rao-Blackwellised particle filter

Fredrik Lindsten, Thomas B. Schön, Jimmy Olsson

Division of Automatic Control
E-mail: lindsten@isy.liu.se, schon@isy.liu.se, jimmy@maths.lth.se

20th October 2010

Report no.: LiTH-ISY-R-2979
Submitted to 18th World Congress of the International Federation of Automatic Control (IFAC)

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se
Abstract

Particle filters (PFs) have shown to be very potent tools for state estimation in nonlinear and/or non-Gaussian state-space models. For certain models, containing a conditionally tractable substructure (typically conditionally linear Gaussian or with finite support), it is possible to exploit this structure in order to obtain more accurate estimates. This has become known as Rao-Blackwellised particle filtering (RBPF). However, since the RBPF is typically more computationally demanding than the standard PF per particle, it is not always beneficial to resort to Rao-Blackwellisation. For the same computational effort, a standard PF with an increased number of particles, which would also increase the accuracy, could be used instead. In this paper, we have analysed the asymptotic variance of the RBPF and provide an explicit expression for the obtained variance reduction. This expression could be used to make an efficient discrimination of when to apply Rao-Blackwellisation, and when not to.

Keywords: Particle filtering, Monte-Carlo methods, Rao-Blackwellised particle filter, Marginalised particle filter, Rao-Blackwellisation, variance reduction
An explicit variance reduction expression for the Rao-Blackwellised particle filter

Fredrik Lindsten * Thomas B. Schön * Jimmy Olsson **

* Division of Automatic Control, Linköping University, Linköping, Sweden (e-mail: {lindsten, schon}@isy.liu.se)
** Center of Mathematical Sciences, Lund University, Lund, Sweden (e-mail: jimmy@maths.lth.se)

Abstract: Particle filters (PFs) have shown to be very potent tools for state estimation in nonlinear and/or non-Gaussian state-space models. For certain models, containing a conditionally tractable substructure (typically conditionally linear Gaussian or with finite support), it is possible to exploit this structure in order to obtain more accurate estimates. This has become known as Rao-Blackwellised particle filtering (RBPF). However, since the RBPF is typically more computationally demanding than the standard PF per particle, it is not always beneficial to resort to Rao-Blackwellisation. For the same computational effort, a standard PF with an increased number of particles, which would also increase the accuracy, could be used instead. In this paper, we have analysed the asymptotic variance of the RBPF and provide an explicit expression for the obtained variance reduction. This expression could be used to make an efficient discrimination of when to apply Rao-Blackwellisation, and when not to.

Keywords: Particle filtering, Monte-Carlo methods, Rao-Blackwellised particle filter, Marginalised particle filter, Rao-Blackwellisation, variance reduction

1. INTRODUCTION AND RELATED WORK

Many important problems in various fields of science are related to state estimation in general state-space models, based on noisy observations. If a prior distribution is assumed for the initial state, the optimal filter is given by the Bayesian filtering recursions. In a few special cases, basically for linear Gaussian state-space (LGSS) models and finite state-space (FSS) models, the optimal filtering problem is analytically tractable. However, many interesting problems do not exhibit such nice properties, but are both nonlinear and/or non-Gaussian. In these cases, the optimal filter needs to be approximated in some way. Sequential Monte Carlo methods, or particle filters (PFs), have shown to be very powerful tools when addressing such intractable models. Since the introduction of the PF by Gordon et al. (1993), we have experienced a vast amount of research in the area. For instance, many improvements and extensions have been introduced to increase the accuracy of the filter, see e.g. Doucet and Johansen (2010) for an overview of recent developments.

One natural idea is to exploit any tractable substructure in the model, see e.g. Doucet et al. (2000b); Schön et al. (2005). More precisely, if the model, conditioned on one partition of the state, behaves like e.g. an LGSS or an FSS it is sufficient to employ particles for the intractable part and make use of the analytic tractability for the remaining part. Inspired by the Rao-Blackwell theorem, this has become known as the Rao-Blackwellised particle filter (RBPF).

The motivation for the RBPF is to improve the accuracy of the filter, i.e. any estimator derived from the RBPF will intuitively have lower variance than the corresponding estimator derived from the standard PF. Informally, the reason for this is that in the RBPF, the particles are spread in a lower dimensional space, yielding a denser particle representation of the underlying distribution. The improved accuracy is also something that is experienced by practitioners. However, it can be argued that it is still not beneficial to resort to Rao-Blackwellisation in all cases. The reason is that the RBPF is typically more computationally expensive per iteration, compared to the standard PF (e.g. for an RBPF targeting a conditional LGSS model, each particle is equipped with a Kalman filter, all which needs to be updated at each iteration). Hence, for a fixed computational effort, we can choose to either use Rao-Blackwellisation or to run a standard PF, but instead increase the number of particles. Both these alternatives will reduce the variance of the estimators. Hence, it is important to understand and to be able to quantify how large variance reduction we can expect from the RBPF, in order to make suitable design choices for any given problem.

In this paper we shall study the asymptotic (in the number of particles) variances for the RBPF and the standard PF (we shall throughout the paper use the abbreviation SPF when referring to the non-Rao-Blackwellised PF). We provide an explicit expression for the difference between the variance of an estimator derived from the SPF and the variance of the corresponding estimator derived from...
All random variables are defined on a common probability distribution. 2.1 Notation

2.1 Notation

2.2 Particle filtering

2.2 Particle filtering

2.3 Rao-Blackwellised particle filter

2.3 Rao-Blackwellised particle filter
Algorithm 1: Particle filter

**Input:** A weighted sample \( \{x_{N,i}^{1:t-1}, w_{t-1}^{N,i}\}_{i=1}^N \) targeting \( \phi_{t-1} \).

**Resampling:** Sample \( N \) indices from a discrete distribution, i.e. for \( i = 1, \ldots, N \),

\[
P \left( \tilde{x}_{N,i}^{1:t} = j \right| \{x_{N,k}^{1:t}, w_{t-1}^{N,k}\}_{k=1}^N \} = w_{t-1}^{N,j}/\sum_{i=1}^N w_{t-1}^{N,i}.
\]

Set \( \tilde{x}_{N,i}^{1:t-1} = x_{N,i}^{1:t-1}, i = 1, \ldots, N \). The equally weighted sample \( \{x_{N,i}^{1:t-1}\}_{i=1}^N \) targets \( \phi_{t-1} \).

**Importance sampling:** Choose a proposal kernel according to (2). Sample new particles according to

\[
x_{i}^{N,i,1:t} \sim R_{t-1}(x_{i}^{N,i,1:t-1}, dx_{i,1:t}), \quad i = 1, \ldots, N,
\]

where \( x_{i}^{N,i,1:t} \) identifies to \( x_i \). Here, \( \phi_{i}^{n} \) is the marginal smoothing distribution of \( Z_{1:t} \) and \( \phi_{i}^{t} \) is the conditional smoothing distribution of \( Z_{1:t} \) given \( \Xi_{1:t} = \xi_{1:t} \). The conditional distribution is assumed to be analytically tractable, typically Gaussian or with finite support.

**Remark 1.** More precisely, \( \phi_{i}^{n} \) is a kernel from \( X_{1}^{n} \) to \( X_{1} \). For each fixed \( \xi_{1:t} \), \( \phi_{i}^{n}(\cdot | \xi_{1:t}) \) is a measure on \( X_{1}^{n} \), and can hence be viewed as a conditional distribution. In the notation introduced in (5), the meaning is that \( \phi_{i}^{n} \) is the product of the measure \( \phi_{i}^{m} \) and the kernel \( \phi_{j}^{n} \). In the remainder of this paper we shall make frequent use of a Fubini like theorem for such products, see e.g. Uglanov (1991).

Instead of running the SPF targeting the “full” smoothing distribution (1), we have the option to target the marginal distribution \( \phi_{i}^{m} \) with a PF and then make use of an analytical expression for \( \phi_{i}^{n} \). Hence, we choose a proposal kernel \( R_{t-1}^{m}(\xi_{1:t-1}, dx_{i,1:t}) \) from \( X_{1}^{m} \) to \( X_{1}^{n} \) such that \( \phi_{i}^{n} \ll \pi_{i}^{m} \) and define a weight function \( W_{i}^{m}(\xi_{1:t}) \) analogously to (4). The measure \( \pi_{i}^{m} \) is defined analogously to (3).

A weighted particle system \( \{x_{N,i}^{1:t}, w_{t}^{N,i}\}_{i=1}^N \), targeting \( \phi_{t}^{n} \), can then be generated in the same manner as in Algorithm 1. We simply replace \( \{x_{N,i}^{1:t}, w_{t-1}^{N,i}\}_{i=1}^N \) with \( \{x_{N,i}^{1:t}, w_{t}^{N,i}\}_{i=1}^N \) and \( \phi_{i}, R, W, \phi_{i}^{m}, R_{t}^{m}, W_{t}^{m} \), respectively (again, superscript \( m \) for marginal). This will be referred to as the Rao-Blackwellised particle filter (RBPF).

**Remark 2.** The most common way to present the RBPF is for conditional LGSS models. In this case, the conditional distribution \( \phi_{t}^{n} \) is Gaussian, which means that it can be computed using the Kalman filter recursions. Consequently, the Kalman filter updates are often shown as intrinsic steps in the presentation of the RBPF algorithm, see e.g. Schön et al. (2005). In this paper, we adopt a more general view and simply see the RBPF as a regular PF, targeting the marginal distribution \( \phi_{t}^{m} \). We then assume that the conditional distribution \( \phi_{t}^{n} \) is available by some means (for the conditional LGSS case, this would of course be by the Kalman filter), but it is not important for our results what those means are.

3. PROBLEM FORMULATION

The SPF and the RBPF can both be used to estimate expectations under the smoothing distribution. Assume that we, for some function \( f \in L^{1}(X_{1}^{n}, \phi_{t}) \), seek the expectation \( \phi_{t}(f) \). For the SPF we use the natural estimator,

\[
\hat{f}_{t}^{SPF} \triangleq \sum_{i=1}^{N} \frac{w_{t}^{N,i}}{\sum_{j=1}^{N} w_{t}^{N,j}} f(x_{N,i}^{1:t}). \quad (6)
\]

For the RBPF we use the fact that \( \phi_{t}(f) = \phi_{t}^{m}(\phi_{t}^{n}(f)) \), and define the estimator,

\[
\hat{f}_{t}^{RBPF} = \sum_{i=1}^{N} \frac{w_{t}^{N,i}}{\sum_{j=1}^{N} w_{t}^{N,j}} \phi_{t}^{m} \left( f(\{x_{N,i}^{1:t}\}) \mid \xi_{N,i}^{1:t} \right). \quad (7)
\]

The question then arise, how much better is (7) compared to (6)?

One analysis of this question, sometimes seen in the literature, is to simply consider a decomposition of variance,

\[
\text{Var}(f) = \text{Var}(E[f \mid \Xi_{1:t}]) + E[\text{Var}(f \mid \Xi_{1:t})]. \quad (8)
\]

Here, the last term is claimed to be the variance reduction obtained in the RBPF. The decomposition is of course valid, the problem is that it does not answer our question. What we have in (8) is simply an expression for the variance of the test function \( f \), it does not apply to the estimators (6) and (7).

**Remark 3.** It is not hard to see why the “simplified” analysis (8) has been considered. If the PF would produce independent and identically distributed (i.i.d.) samples from the target distribution (which it does not), then the analysis would be correct. More precisely, for i.i.d. samples, the central limit theorem states that the asymptotic variance of an estimator of a test function \( f \), coincides with the variance of the test function itself (up to a factor \( 1/N \)). However, as we have already pointed out, the PF does not produce i.i.d. samples. This is due to the resampling step, in which a dependence between the particles is imposed. At the end of Section 6, one of the inadequacies of (8) will be pointed out.

Hence, we are interested in the asymptotic variance of (6) and (7), respectively. To analyse this we shall borrow the concept of asymptotic normality from Douc and Moulines (2008).

**Definition 1.** (Asymptotic normality) Let \((X, \mathcal{X})\) be a measurable space, \(A\) and \(W\) subsets of \(\mathcal{F}(X)\), \(\mu\) a probability measure and \(\gamma\) a finite measure, both on \((X, \mathcal{X})\). Let \(\sigma\) be a real nonnegative function on \(A\) and \(\{\alpha_N\}_{N=1}^{\infty}\) a nondecreasing real sequence diverging to infinity.

A weighted sample \(\{x_{N,i}^{1:t}, w_{t}^{N,i}\}_{i=1}^{N}\) is said to be asymptotically normal for \((\mu, A, W, \sigma, \gamma, \{\alpha_N\})\) if

\[
a_{N}\alpha_{N}^{-1} \sum_{i=1}^{N} w_{t}^{N,i} \left( f(x_{N,i}^{1:t}) - \mu(f) \right) \overset{D}{\to} N(0, \sigma^2(f)), \quad (9)
\]
\[
a_{\xi} N^{-\frac{1}{2}} \max_{1 \leq i \leq N} u_{N,i} \xrightarrow{p} 0, \tag{11}
\]
as \( N \to \infty \), for any \( f \in \mathcal{A} \) and any \( g \in \mathcal{W} \), where \( \Omega_N = \sum_{j=1}^{N} v_{N,j} \).

In the following two theorems (slight modifications of what has previously been given by Douc and Moulines (2008)) we claim asymptotic normality for the weighted particle systems generated by the SPF and the RBPF, respectively.

**Theorem 1.** (Asymptotic normality of the SPF) Assume that the initial particle system \( \{x_{N,i}^{t}, w_{N,i}^{t}\}_{i=1}^{N} \) is asymptotically normal for \( \langle \phi_1, A_1, W_1, \sigma_1, \phi_1, \{\sqrt{N}\} \rangle \). Define recursively the sets
\[
A_t = \{ f \in L^2(\mathcal{X}^t, \phi_t) : R_{t-1}(\cdot, W_{t-1} f) \in \mathcal{A}_{t-1}, \quad R_{t-1}(\cdot, W_{t-1} f^2) \in \mathcal{W}_{t-1} \},
\]
\[
W_t = \{ f \in L^1(\mathcal{X}^t, \phi_t) : R_{t-1}(\cdot, W_{t-1} f) \} \in \mathcal{W}_{t-1} \}.
\]
Assume that, for any \( t \geq 1 \), \( R_t(\cdot, W_t^2) \in \mathcal{W} \). Then, for any \( t \geq 1 \), the weighted particle system \( \{x_{N,i}^{t}, w_{N,i}^{t}\}_{i=1}^{N} \) generated by the SPF is asymptotically normal for \( \langle \phi_t, A_t, W_t, \sigma_t, \phi_t, \{\sqrt{N}\} \rangle \). The asymptotic variance is, for \( f \in A_{t-1} \), given by
\[
\sigma^2_t(f) = \sigma^2_{t-1}(R_{t-1}(\cdot, W_{t-1} f)) + \phi_{t-1}[R_{t-1}(\cdot, (W_{t-1} f)^2)], \tag{12}
\]
\[
\tilde{f} = f - \phi_t(f).
\]
**Proof.** See Appendix A.

**Theorem 2.** (Asymptotic normality of the RBPF) Under analogous conditions and definitions as in Theorem 1, for any \( t \geq 1 \) the particle system \( \{\xi_{1:i}^t, \omega_{1:i}^t\}_{i=1}^N \) generated by the RBPF is asymptotically normal for \( \{\phi^m_{t}, A^m_t, W^m_t, \tau^m_t, \phi^m_t, \{\sqrt{N}\} \rangle \). The asymptotic variance is, for \( g \in A^m_{t-1} \), given by
\[
\tau^2_t(g) = \tau^2_{t-1}(R^m_{t-1}(\cdot, W^m_{t-1} g)) + \phi^m_{t-1}(R^m_{t-1}(\cdot, (W^m_{t-1} g)^2)), \tag{13}
\]
\[
g = g - \phi^m_t(g).
\]
**Proof.** See Appendix A.

Recall from Remark 2 that the SPF and the RBPF are really just two particle filters, targeting different distributions, hence the similarity between the two theorems above. Actually, we could have sufficed with one, more general, theorem applicable to both filters. The reason for why we have chosen to present them separately is for clarity and to introduce all the required notation.

As previously pointed out, the RBPF will intuitively produce better estimates than the SPF, i.e. we expect \( \tau^2_t(\phi_t(f)) \leq \sigma^2_t(f) \). Let us therefore define the variance difference
\[
\Delta_t(f) \triangleq \sigma^2_t(f) - \tau^2_t(\phi_t(f)). \tag{14}
\]
The problem that we are concerned with in this paper is to find an explicit expression for this quantity. This will be provided in the next section.

### 4. THE MAIN RESULT

To analyse the variance difference (14) we shall need the following assumption (similar to what is used by Chopin (2004)).

**Assumption 1.** For each \( \tilde{\xi}_{1:t-1} \in \mathcal{X}^{t-1}_\xi \), the two measures
\[
\int_{\mathcal{X}^{t-1}_\xi} \phi_{t-1}^*(d\tilde{\xi}_{1:t-1} | \tilde{\xi}_{1:t-1}) R_{t-1}(\{\tilde{\xi}_{1:t-1}, \tilde{z}_{1:t-1}\}, dx_{1:t}) \tag{15}
\]
and
\[
a_t(\xi_{1:t}) R^m_{t-1}(\tilde{\xi}_{1:t-1}, d\xi_{1:t}) \sigma^m_t(\tilde{d}z_{1:t} | \xi_{1:t}) \tag{16}
\]
agree on \( \mathcal{X}^t \), for some positive function \( a_t : \mathcal{X}^t_{\xi} \to \mathbb{R} \) and some transition kernel \( \sigma^m_t \) from \( \mathcal{X}^t_{\xi} \) to \( \mathcal{X}^t_{\xi} \), for which \( \phi_t^*(\cdot | \xi_{1:t}) \ll \sigma^m_t(\cdot | \xi_{1:t}) \).

The basic meaning of this assumption is to create a connection between the proposal kernels \( R_{t-1} \) and \( R^m_{t-1} \). It is natural that we need some kind of connection. Otherwise the asymptotic variance expressions (12) and (13) would be completely decoupled, and it would not be possible to draw any conclusions from a comparison. Still, as we shall see in the next section, Assumption 1 is fairly weak.

We are now ready to state the main result of this paper.

**Theorem 3.** Under Assumption 1, and using the definitions from Theorem 1 and Theorem 2, for any \( f \in \mathcal{A}_t \),
\[
\Delta_t(f) = \Delta_{t-1}(R_{t-1}(\cdot, W_{t-1} f)) + \phi_{t-1}^m \left[ R^m_{t-1}(\cdot, (W^m_{t-1} \tilde{f})^2) + a_t \var_{t}(W_{t-1} f) \right], \tag{17}
\]
where
\[
\bar{\Psi} = \phi_t(f) - \phi_t(f),
\]
\[
\tilde{A}_t \triangleq \{ f \in F(\mathcal{X}^t) : \phi_t(f) \in \mathcal{A}^m_t \cap \mathcal{A}_t \}. \tag{19}
\]
**Proof.** See Appendix A.

### 5. RELATIONSHIP BETWEEN THE PROPOSALS KERNELS

To understand the results given in the previous section, we shall have a closer look at the relationship between the proposal kernels induced by Assumption 1. We shall do this for a certain family of proposal kernels. More precisely, assume that the kernels can be written
\[
R_{t-1}(x_{1:t-1}, dx_{1:t}) = r_{t-1}(dx_{1:t} | x_{1:t-1}) \delta_{\xi_{1:t-1}}(dx_{1:t-1}), \tag{20}
\]
\[
R^m_{t-1}(\tilde{\xi}_{1:t-1}, d\xi_{1:t}) = r^m_{t-1}(d\xi_{1:t} | \xi_{1:t-1}) \delta_{\tilde{\xi}_{1:t-1}}(d\xi_{1:t-1}). \tag{21}
\]
Informally, this means that when a trajectory \( \{x^N_{1:i} \} \) is sampled at time \( t \), we keep the “old” trajectory up to time \( t-1 \) and simply append a sample from time \( t \). This is the case for most PFs (when targeting the joint smoothing distribution), but not all, see e.g. the resample-move algorithm by Gilks and Berzuini (2001).

Furthermore, let \( r_{t-1} \) be factorised as
\[
r_{t-1}(dx_{1:t} | x_{1:t-1}) = q_{t-1}^m(d\xi_{1:t} | x_{1:t-1}) q_{t-1}^m(d\xi_{1:t} | x_{1:t-1}, z_{1:t-1}), \tag{22}
\]
Assume that $q^m_{t-1} \leq r^m_{t-1}$ and define the kernel
\[
\nu_t(d z_{1:t} \mid \xi_{1:t}) \triangleq \frac{d q^m_{t-1} \cdot (\xi_{1:t-1}, z_{1:t-1})}{d r^m_{t-1} \cdot (\xi_{1:t-1})} \times \phi^c_{t-1}(d z_{1:t-1} \mid \xi_{1:t-1}) q^m_{t-1}(dz_t \mid \xi_{1:t}, z_{1:t-1}).
\] (23)

It can now be verified that the choice
\[
a_t(\xi_{1:t}) = \int_{\mathcal{Z}_t} \nu_t(d z_{1:t} \mid \xi_{1:t}),
\] (24)

\[
\pi^c_t(d z_{1:t} \mid \xi_{1:t}) = \frac{\nu_t(d z_{1:t} \mid \xi_{1:t})}{a_t(\xi_{1:t})},
\] (25)
satisfies Assumption 1, given that $\phi^c_t(\cdot \mid \xi_{1:t}) \ll \pi^c_t(\cdot \mid \xi_{1:t})$.

Hence, the function $a_t$ takes the role of a normalisation of the kernel $\nu_t$ to obtain a transition kernel $\pi^c_t$. One interesting fact is that, from (17), we cannot guarantee that $\Delta_t(f)$ is nonnegative for arbitrary functions $a_t$. At first this might seem counterintuitive, since it would mean that the variance is higher for the RBPF than for the SPF. The explanation lies in that Assumption 1, relating the proposal kernels in the two filters, is fairly weak. In other words, we have not assumed that the proposal kernels are “equally good”. For instance, say that the optimal proposal kernel is used in the SPF, whereas the RBPF uses a poor kernel. It is then no longer clear that the RBPF will outperform the SPF. In the next section we shall see that if both filters use their respective bootstrap proposal kernel, a case when the term “equally good” makes sense, then $\Delta_t(f)$ will indeed be nonnegative. However, for other proposal kernels, it is not clear that there is an analogue between the SPF and the RBPF in the same sense.

5.1 Example: Bootstrap kernels

Let $Q(dx_t \mid x_{1:t-1})$ be the Markov transition kernel of the process $X$. In the bootstrap SPF we choose the proposal kernel according to (20) with
\[
r^m_{t-1}(dx_t \mid x_{1:t-1}) = Q(dx_t \mid x_{1:t-1}),
\] (26)

where, almost surely, for $A \in \mathcal{X}$
\[
Q(A \mid X_{t-1}) = P(X_t \in A \mid X_{t-1}) = P(X_t \in A \mid X_{1:t-1}, Y_{1:t-1}).
\] (27)

The second equality follows from the Markov property of the process. In the RBPF, the analogue of the bootstrap proposal kernel is to use (21) with
\[
r^m_{t-1}(A \mid \Xi_{1:t-1}) = P(\Xi_t \in A \mid \Xi_{1:t-1}, Y_{1:t-1}),
\] (28)

for $A \in \mathcal{X}_t$.

It can be verified (see Appendix B) that these choices fulfill Assumption 1 with
\[
a_t \equiv 1,
\] (29)

and
\[
\pi^c_t(A \mid \Xi_{1:t}) = P(Z_{1:t} \in A \mid \Xi_{1:t}, Y_{1:t-1}),
\] (30)

for $A \in \mathcal{X}_t$. Hence, $\pi^c_t$ is indeed the predictive distribution of $Z_{1:t}$ conditioned on $\Xi_{1:t}$ and based on the measurements up to time $t-1$. In this case we can also write $\pi^c_t(dx_{1:t}) = \pi^m_t(dx_{1:t}) \pi^c_t(z_{1:t-1} \mid \xi_{1:t})$, which highlights the connection between the predictive distributions in the two filters. In this case, due to (29), the variance difference (17) can be simplified to
\[
\Delta_t(f) = \Delta_{t-1}(R_{t-1}(\cdot, W_{t-1}f)) + \phi^m_{t-1}[R_{t-1}(\cdot, \text{Var}_t(W_{t-1}f))].
\] (31)

Hence, $\Delta_t(f)$ can be written as a sum (though, we have expressed it in a recursive form here) in which each term is an expectation of a conditional variance. It is thus ensured to be nonnegative.

6. DISCUSSION

In Theorem 3 we gave an explicit expression for the difference in asymptotic variance between the SPF and the RBPF. This expression can be used as a guideline for when it is beneficial to apply Rao-Blackwellisation, and when it is not. The variance expressions given in this paper are asymptotic. Consequently, they do not apply exactly to the variances of the estimators (6) and (7), for a finite number of particles. Still, a reasonable approximation of the accuracy of the estimator (6) is
\[
\text{Var} \left( \hat{f}^N_{\text{SPF}} \right) \approx \frac{\sigma^2(f)}{N},
\] (32)

and similarly for (7)
\[
\text{Var} \left( \hat{f}^N_{\text{RBPF}} \right) \approx \frac{\tau^2(\phi^c(f))}{N}.
\] (33)

Now, assume that the computational effort required by the RBPF, using $M$ particles, equals that required by the SPF, using $N$ particles (thus, $M < N$ since, in general, the RBPF is more computationally demanding than the SPF per particle). We then have
\[
\frac{\text{Var} \left( \hat{f}^N_{\text{SPF}} \right)}{\text{Var} \left( \hat{f}^M_{\text{RBPF}} \right)} \approx \frac{N}{M} \left( 1 + \frac{\Delta_t(f)}{\tau_t^2(\phi^c(f))} \right).
\] (34)

Whether or not this quantity is greater than one tells us if it is beneficial to use Rao-Blackwellisation. The crucial point is then to compute the ratio $\Delta_t(f)/\tau_t^2(\phi^c(f))$, which in itself is a challenging problem. We are currently investigating ways to estimate this ratio from a single run of the RBPF.

As a final remark, for the special case discussed in Section 5.1, the variance difference (31) resembles the last term in the expression (8). They are both composed of an expectation of a conditional variance. One important difference though, is that the dependence on the weight function $W_{t-1}$ is visible in (31). As an example, if the test function is restricted to $f \in L^1(X_t, \phi^m_t)$ the gain in variance indicated by (8) would be zero (since $\text{Var}(f(\Xi_{1:t}) \mid \Xi_{1:t}) \equiv 0$), but this is not the case for the actual gain (31).

7. CONCLUSIONS

We have analysed the Rao-Blackwellised particle filter in a fairly general setting, and provide an explicit expression for the reduction of asymptotic variance obtained from Rao-Blackwellisation. This expression is expected to be of practical use, since it can serves as an indicator for when it is beneficial to apply Rao-Blackwellisation, and when it is not. We are currently investigating efficient methods, based on the analytical expression, for estimating the obtained variance reduction.

Appendix A. PROOFS

Proof of Theorem 1. In Theorem 10 in Douc and Moulines (2008), take
Proof of Theorem 2. As the previous proof, with \( \kappa \) which satisfies the conditions of the hypothesis. Further-
more, \( \phi_{t-1}(\ldots) = 1 \). Now, the results follow for the choice \( \kappa = 0 \).

In Theorem 10 in Douc and Mounines (2008), the asymptotic normality of a particle system obtained after resampling is considered. Compared to Theorem 1 of this paper, they thus obtain an additional term in the expression for the asymptotic variance.

Proof of Theorem 2. As the previous proof, with

\[
L_{t-1}^m(\xi_{1:t-1}, \xi_{1:t-1}) = W_{t-1}(\xi_{1:t-1})R_{t-1}(\xi_{1:t-1}, \xi_{1:t-1}).
\]

Proof of Theorem 3. Let Assumption 1 be satisfied. We shall start by determining the relationship between the weight functions \( W_{t-1} \) and \( W_{t-1}^m \). Consider

\[
\phi_t(dx_{1:t}) = \frac{d\phi_t}{d\pi_t}(x_{1:t})\pi_t(dx_{1:t})
\]

\[
= W_{t-1}(x_{1:t}) \int_{\chi_{t-1}} \phi_{t-1}(d\tilde{x}_{1:t-1})R_{t-1}(\tilde{x}_{1:t-1}, dx_{1:t})
\]

where we have made use of the definitions in (3) and (4). Furthermore, from the factorisation of \( \phi_{t-1} \) and Assumption 1 we get

\[
\phi_t(dx_{1:t}) = W_{t-1}(x_{1:t}) \int_{\chi_{t-1}} \phi_{t-1}(d\tilde{x}_{1:t-1})R_{t-1}(\tilde{x}_{1:t-1}, dx_{1:t})
\]

where \( \phi_{t-1}(d\tilde{x}_{1:t-1}) \) is the Radon-Nikodym derivative of \( \phi_{t-1} \) with respect to \( \pi_{t-1} \). Let \( \sigma^m_{t-1}(\cdot) \) denote the \( \pi_{t-1}^m \)-almost surely

\[
\pi_t^m(W_{t-1}\tilde{f}) = \int \frac{W_{t-1}(\xi_{1:t})}{a_t(\xi_{1:t})} \frac{d\phi_t^m(\cdot | \xi_{1:t})}{d\pi_t^m(\cdot | \xi_{1:t})} (z_{1:t}) \times f(\xi_{1:t}, z_{1:t}) \pi_t^m(z_{1:t} | \xi_{1:t}),
\]

(\text{A.8})

Here we have used the definition of \( \tilde{f} \) in (18), yielding

\[
\phi_t^m(\tilde{f}) = \phi_t^m(f - \phi_t(f)) = \tilde{f}(\xi_{1:t}).
\]

(A.9)

Combining (A.7) and (A.8) we get, \( \pi_t^m \)-almost surely,

\[
\begin{align*}
\pi_t^m(W_{t-1}\tilde{f}) & = \frac{(W_{t-1}(\xi_{1:t})\tilde{f}(\xi_{1:t}))^2}{a_t(\xi_{1:t})} + \pi_t^m(\tilde{f}) Var_{\pi_t^m}(W_{t-1}\tilde{f}).
\end{align*}
\]

(\text{A.10})

Let \( L_{t-1} \) and \( L_{t-1}^m \) be defined as in (A.1) and (A.2), respectively. Then,

\[
\Delta_t(f) = \sigma^2_{t-1}(L_{t-1}(\cdot, f)) - \sigma^2_{t-1}(L_{t-1}(\cdot, \tilde{f})).
\]

Using (14), (12), (13) and the above results, the difference in asymptotic variance can be expressed as

\[
\begin{align*}
\Delta_t(f) & = \pi_t^m(L_{t-1}(\cdot, f)) - \pi_t^m(L_{t-1}(\cdot, \tilde{f})).
\end{align*}
\]

(\text{A.11})

(recall that \( \pi_t^m = \pi_t^m R_{t-1}^m \), which ensures that we, due to the expectation w.r.t. \( \phi_{t-1} R_{t-1}^m \) in (A.11), can make use of the equality in (A.10)).

Finally, consider

\[
\phi_{t-1}(L_{t-1}(\cdot, f)) = \int_{\chi_t} \left( \int_{\chi_t} \phi_t(dx_{1:t}) \bar{f}(x_{1:t}) \times \int_{\chi_{t-1}} \phi_{t-1}(d\tilde{x}_{1:t-1} | \tilde{x}_{1:t-1})R_{t-1}(\tilde{x}_{1:t-1}, dx_{1:t}) \right)
\]

\[
\begin{align*}
& = \int_{\chi_t} \int_{\chi_{t-1}} \phi_t(dx_{1:t}) W_{t-1}(\xi_{1:t}, z_{1:t}) \pi_t^m(dz_{1:t} | \xi_{1:t})
\end{align*}
\]

(\text{A.12})

Hence, consider

\[
\begin{align*}
\phi_{t-1}(L_{t-1}(\cdot, f)) & = \phi_{t-1}^m \left[ R_{t-1}(\cdot, a_t(\cdot, \pi_t^m(\cdot))) \right].
\end{align*}
\]

(\text{A.6})

Comparing (A.6) and (A.2), we see that we can let \( f \) take the role of \( (W_{t-1}\tilde{f})^2 \). Hence, consider

\[
\begin{align*}
\pi_t^m((W_{t-1}\tilde{f})^2) & = \left( \pi_t^m(W_{t-1}\tilde{f}) \right)^2 + Var_{\pi_t^m}(W_{t-1}\tilde{f}),
\end{align*}
\]

(\text{A.7})

where, using (A.5) we have \( \pi_t^m \)-almost surely,
Appendix B. VERIFICATION OF ASSUMPTION 1
FOR BOOTSTRAP KERNELS

We shall now analyse Assumption 1 for the choice of proposal kernels used in Section 5.1. First, note that, for $A \in \mathcal{X}_t$

$$\phi_t^\mu(A \mid \Xi_{1:t}) = P(Z_{1:t} \in A \mid \Xi_{1:t}, Y_{1:t}). \quad (B.1)$$

Hence, the two measures agree on measurable rectangles. Now, let $\mu^1(\cdot \mid \xi_{1:t-1})$ be the measure in (15). Let $\{A, B, C, D\}$ be a measurable rectangle, i.e. $A \in \mathcal{X}_t$, $B \in \mathcal{X}_t$, $C \in \mathcal{X}_t^{-1}$, $D \in \mathcal{X}_t^{-1}$, and consider

$$\mu^1(A \times B \times C \times D) \mid \Xi_{1:t-1}$$

$$= \int_A \int_B \int_C \int_D \phi_t^{-1}(d\tilde{z}_{1:t-1} \mid \Xi_{1:t-1})$$

$$\times r_{t-1}(d\{\xi_t, z_t\} \mid \{\xi_{1:t-1}, z_{1:t-1}\})$$

$$\times \delta_{z_{1:t-1}}(d\xi_{1:t-1}, d\tilde{z}_{1:t-1})$$

$$= \int_A \int_B \int_C \int_D \phi_t^{-1}(d\tilde{z}_{1:t-1} \mid \Xi_{1:t-1})$$

$$\times r_{t-1}(d\{\xi_t, z_t\} \mid \{\xi_{1:t-1}, z_{1:t-1}\})$$

$$\times \delta_{z_{1:t-1}}(d\xi_{1:t-1}, d\tilde{z}_{1:t-1})$$

$$= I_C(\Xi_{1:t-1}) \int_A \int_B \int_D \phi_t^{-1}(d\tilde{z}_{1:t-1} \mid \Xi_{1:t-1})$$

$$\times r_{t-1}(d\{\xi_t, z_t\} \mid \{\xi_{1:t-1}, z_{1:t-1}\})$$

$$\times \delta_{z_{1:t-1}}(d\xi_{1:t-1}, d\tilde{z}_{1:t-1})$$

$$= I_C(\Xi_{1:t-1}) \int_A \int_B \int_D r_{t-1}(d\{\xi_t, z_t\} \mid \{\xi_{1:t-1}, Z_{1:t-1}\})$$

$$\times \delta_{z_{1:t-1}}(d\xi_{1:t-1}, d\tilde{z}_{1:t-1})$$

/using (26) and (27)/

$$= I_C(\Xi_{1:t-1}) \int_B \int_D I_D(Z_{1:t-1}) \times$$

$$E[I_A(\Xi_t) I_B(Z_t) \mid \Xi_{1:t-1}, Z_{1:t-1}, Y_{1:t-1}] \mid \Xi_{1:t-1}, Y_{1:t-1}]$$

$$= I_C(\Xi_{1:t-1}) \int_B \int_D I_D(Z_{1:t-1}) I_A(Z_t) I_B(Z_t) \mid \Xi_{1:t-1}, Y_{1:t-1}]$$

Similarly, let $\mu^2(\cdot \mid \tilde{\xi}_{1:t-1})$ be the measure (16). We shall see that the two measures $\mu^1$ and $\mu^2$ agree if (29) and (30) holds true. Consider

$$\mu^2(A \times B \times C \times D) \mid \Xi_{1:t-1}$$

$$= \int_A \int_B \int_C \int_D \nabla_t^r(dz_{1:t} \mid \xi_{1:t})$$

$$\times r_{t-1}^m(d\xi_{1:t-1}) \delta_{z_{1:t-1}}(d\xi_{1:t-1})$$

$$= I_C(\Xi_{1:t-1})$$

$$\times \int_A \int_B \int_D \nabla_t^r(dz_{1:t} \mid \Xi_{1:t-1}, \xi_t)$$

$$\times r_{t-1}^m(d\xi_{1:t-1}) \delta_{z_{1:t-1}}(d\xi_{1:t-1})$$

/using (28)/

$$= I_C(\Xi_{1:t-1})$$

$$\times E[I_A(\Xi_t) \int_B \int_D \nabla_t^r(dz_{1:t} \mid \Xi_{1:t}) \mid \Xi_{1:t-1}, Y_{1:t-1}]$$

/integrating over $z_{1:t}$, using (30)/

$$= I_C(\Xi_{1:t-1}) E[I_A(\Xi_t) E[I_D(Z_{1:t-1}) I_D(Z_t) \mid \Xi_{1:t-1}, Y_{1:t-1}]$$

$$\mid \Xi_{1:t-1}, Y_{1:t-1}]$$

$$= I_C(\Xi_{1:t-1}) E[I_A(\Xi_t) I_D(Z_{1:t-1}) I_D(Z_t) \mid \Xi_{1:t-1}, Y_{1:t-1}]$$.

Hence, the two measures agree on measurable rectangles. Since the measurable rectangles form a $\pi$-system, genera-