

FURTHER RAO-BLACKWELLIZING AN ALREADY RAO-BLACKWELLIZED ALGORITHM FOR JUMP MARKOV STATE SPACE SYSTEMS

Yohan Petetin, François Desbouvries

Telecom Institute / Telecom SudParis
CITI Department and CNRS UMR 5157
9 rue Charles Fourier, 91011 Evry, France

{yohan.petetin, francois.desbouvries}@telecom-sudparis.eu

ABSTRACT

Exact Bayesian filtering is impossible in Jump Markov State Space Systems (JMSS), even in the simple linear and Gaussian case. Suboptimal solutions include sequential Monte-Carlo (SMC) algorithms which are indeed popular, and are declined in different versions according to the JMSS considered. In particular, Jump Markov Linear Systems (JMLS) are particular JMSS for which a Rao-Blackwellized (RB) Particle Filter (PF) has been derived. The RBPF solution relies on a combination of PF and Kalman Filtering (KF), and RBPF-based moment estimators outperform purely SMC-based ones when the number of samples tends to infinity. In this paper, we show that it is possible to derive a new RBPF solution, which implements a further RB step in the already RBPF with optimal importance distribution (ID). The new RBPF-based moment estimator outperforms the classical RBPF one whatever the number of particles, at the expense of a reasonable extra computational cost.

1. INTRODUCTION

Let \mathbf{x}_n (resp. \mathbf{y}_n) be a sequence of hidden states (of observations) and let us consider the so called JMSS which can be seen as a hidden Markov chain (HMC) in which the transition and likelihood pdfs depend on a realization of a discrete Markov chain (MC) r_n :

$$p(\mathbf{r}_{0:n}, \mathbf{x}_{0:n}, \mathbf{y}_{0:n}) = p(r_0) \prod_{k=1}^n p(r_k | r_{k-1}) \times p(\mathbf{x}_0) \prod_{k=1}^n f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1}, r_k) \prod_{k=0}^n g_k(\mathbf{y}_k | \mathbf{x}_k, r_k) \quad (1)$$

in which $p(\mathbf{x}_k)$, say, is the pdf (w.r.t. Lebesgue measure) of \mathbf{x}_k , $\mathbf{x}_{0:n} = \{\mathbf{x}_k\}_{k=0}^n$, $f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1}, r_k)$ is the transition pdf of MC $\{\mathbf{x}_k\}_{k \geq 0}$ given $\mathbf{r}_{0:k}$, and $g_k(\mathbf{y}_k | \mathbf{x}_k, r_k)$ is the likelihood of \mathbf{y}_k given \mathbf{x}_k and r_k . In a JMSS exact Bayesian filtering is either impossible (in the general case) or an NP-hard problem (in the linear and Gaussian case), so one has to use suboptimal techniques [1] [2] [3]. Among them, SMC methods can be divided into two classes:

We would like to thank the French MOD DGA/MRIS for financial support of the Ph.D. of Y. Petetin.

- The first class approximates pdf $p(\mathbf{x}_n, \mathbf{r}_{0:n} | \mathbf{y}_{0:n})$ by a set of weighted samples $\{w_n^i, \mathbf{x}_n^i, \mathbf{r}_{0:n}^i\}_{i=1}^N$, which is propagated using Sequential Importance Sampling (SIS). Let

$$\Theta_n = \int \phi(\mathbf{x}_n) p(\mathbf{x}_n | \mathbf{y}_{0:n}) d\mathbf{x}_n \quad (2)$$

be a moment of interest. Then one can deduce from $\{w_n^i, \mathbf{x}_n^i, \mathbf{r}_{0:n}^i\}_{i=1}^N$ the following Monte Carlo estimate of Θ_n (CL calls for classical):

$$\hat{\Theta}_n^{\text{CL}} = \sum_{i=1}^N w_n^i \phi(\mathbf{x}_n^i). \quad (3)$$

- The second class relies on the assumption that expectation

$$\mathbb{E}_{p(\mathbf{x}_n | \mathbf{r}_{0:n}, \mathbf{y}_{0:n})}(\phi(\mathbf{x}_n)) = \Phi(\mathbf{r}_{0:n}) \quad (4)$$

is computable. According to Bayes rule,

$$p(\mathbf{x}_n, \mathbf{r}_{0:n} | \mathbf{y}_{0:n}) = p(\mathbf{r}_{0:n} | \mathbf{y}_{0:n}) \times p(\mathbf{x}_n | \mathbf{r}_{0:n}, \mathbf{y}_{0:n})$$

so only $p(\mathbf{r}_{0:n} | \mathbf{y}_{0:n})$ is approximated by a set of weighted samples. This solution is known as RBPF [4] [5] [6]. An estimator of Θ_n is given by

$$\hat{\Theta}_n^{\text{RB}} = \sum_{i=1}^N w_n^i \Phi(\mathbf{r}_{0:n}^i). \quad (5)$$

One can deduce from an asymptotical analysis and from the Rao-Blackwell theorem that estimator $\hat{\Theta}_n^{\text{RB}}$ outperforms the classical one $\hat{\Theta}_n^{\text{CL}}$ [4] [7]. The RBPF is of practical interest in JMLS, i.e in JMSS where transitions (resp. likelihoods) $f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1}, r_k)$ (resp. $g_k(\mathbf{y}_k | \mathbf{x}_k, r_k)$) are Gaussian and linear in \mathbf{x}_{k-1} (resp. \mathbf{x}_k). In such models, $p(\mathbf{x}_n | \mathbf{r}_{0:n}, \mathbf{y}_{0:n})$ is computed by KF so $\hat{\Theta}_n^{\text{RB}}$ could be derived for a large class of functions of interest ϕ .

In this paper we derive a new RBPF algorithm for JMLS. More precisely, we start from the RBPF with optimal ID and show that a further RB step is indeed possible, at a reasonable extra computational cost. As a consequence, the induced moment estimator outperforms $\hat{\Theta}_n^{\text{RB}}$, whatever the number of particles N (i.e, our results are

not asymptotical). We next propose an extension of our algorithm to nonlinear JMSS.

The paper is organized as follows. In section 2, we recall the classical RBPF for JMLS, then we derive our new RB estimator in section 3. The variance properties and computational cost of both estimators are compared in section 4. In section 5 we adapt our algorithms to JMSS which are no longer conditionally linear and Gaussian. We finally run simulations in section 6.

2. THE CLASSICAL RBPF FOR JMLS

Let us consider the following model:

$$r_n \text{ is a discrete MC} \quad (6)$$

$$\mathbf{x}_n = \mathbf{F}_n(r_n)\mathbf{x}_{n-1} + \mathbf{G}_n(r_n)\mathbf{u}_n \quad (7)$$

$$\mathbf{y}_n = \mathbf{H}_n(r_n)\mathbf{x}_n + \mathbf{L}_n(r_n)\mathbf{v}_n \quad (8)$$

where $\mathbf{x}_0, \mathbf{u}_1, \dots, \mathbf{u}_n$ and $\mathbf{v}_0, \dots, \mathbf{v}_n$ are independent and independent of r_0, \dots, r_n . We set $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$, $\mathbf{u}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_n)$ and $\mathbf{v}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_n^v)$. So model (6)-(8) is nothing but a classical linear and Gaussian state-space system, except that its dynamics (given by matrices $\mathbf{F}_n(r_n), \mathbf{G}_n(r_n), \mathbf{H}_n(r_n)$ and $\mathbf{L}_n(r_n)$) depends on the realization of a discrete MC r_n .

As is well known [8] [9], the exact computation of the filtering pdf is not possible in such models because

$$p(\mathbf{x}_n | \mathbf{y}_{0:n}) = \sum_{\mathbf{r}_{0:n}} p(\mathbf{r}_{0:n} | \mathbf{y}_{0:n}) p(\mathbf{x}_n | \mathbf{r}_{0:n}, \mathbf{y}_{0:n})$$

is a Gaussian mixture which grows exponentially with time n . To cope with this issue, suboptimal solutions have been developed. The so called RBPF solutions propagate a set of weighted samples $\{w_n^i, \mathbf{r}_{0:n}^i\}_{i=1}^N$ but assume that $E(\phi(\mathbf{x}_n) | \mathbf{r}_{0:n}, \mathbf{y}_{0:n})$ is computable [4]. We now recall their principle.

As in SMC algorithms, the problem consists in propagating recursively the set $\{w_n^i, \mathbf{r}_{0:n}^i\}_{i=1}^N$. According to (5) we should also compute the moments

$$\{E(\phi(\mathbf{x}_n) | \mathbf{r}_{0:n}^i, \mathbf{y}_{0:n})\}_{i=1}^N = \{\Phi(\mathbf{r}_{0:n}^i)\}_{i=1}^N. \quad (9)$$

This is achieved by using SIS techniques and the KF. Let us assume that at time $n-1$ we have a set of weighted particles $\{w_{n-1}^i, \mathbf{r}_{0:n-1}^i\}_{i=1}^N$, with $\mathbf{r}_{0:n-1}^i \sim q(\mathbf{r}_{0:n-1})$, that approximates $p(\mathbf{r}_{0:n-1} | \mathbf{y}_{0:n-1})$. Then one updates this set of weighted particles as follows:

- Sample $r_n^i \sim q(r_n | \mathbf{r}_{0:n-1}^i)$;
- Compute

$$w_n^i \propto w_{n-1}^i \frac{p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n}^i) p(r_n^i | r_{n-1}^i)}{q(r_n^i | \mathbf{r}_{0:n-1}^i)},$$

$$\text{with } \sum_{i=1}^N w_n^i = 1.$$

In order to prevent the concentration of the total mass on few particles, one can resample trajectories $\{\mathbf{r}_{0:n}^i\}_{i=1}^N$ according to weights $\{w_n^i\}_{i=1}^N$ then set $w_n^i = 1/N$ [10] [11].

The optimal sampling distribution $q(r_n | \mathbf{r}_{0:n-1}^i)$ which minimizes the variance of the weights given $\{\mathbf{r}_{0:n-1}^i\}_{i=1}^N$ and $\mathbf{y}_{0:n}$ is

$$p(r_n | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n}) = \frac{p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i, r_n) p(r_n | r_{n-1}^i)}{p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i)}. \quad (10)$$

Next, given $\mathbf{r}_{0:n}$, model (6)-(8) is linear and Gaussian. So let

$$p(\mathbf{x}_{n-1} | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n-1}) = \mathcal{N}(\mathbf{x}_{n-1}; \mathbf{m}_{n-1|n-1}^i; \mathbf{P}_{n-1|n-1}^i),$$

where $\mathcal{N}(\mathbf{x}; \mathbf{m}; \mathbf{P})$ denotes the Gaussian pdf with variable \mathbf{x} , mean \mathbf{m} and covariance \mathbf{P} . Then $p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i, r_n)$ is a Gaussian pdf whose moments are given by the predicted observation mean and predicted observation covariance of the KF, i.e

$$p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i, r_n) = \mathcal{N}(\mathbf{y}_n; \tilde{\mathbf{y}}_n^i(r_n); \mathbf{S}_n^i(r_n)),$$

where

$$\tilde{\mathbf{y}}_n^i(r_n) = \mathbf{y}_n - \mathbf{H}_n(r_n)\mathbf{m}_{n|n-1}^i(r_n), \quad (11)$$

$$\mathbf{S}_n^i(r_n) = \mathbf{H}_n(r_n)\mathbf{P}_{n|n-1}^i(r_n)\mathbf{H}_n(r_n)^T + \mathbf{L}_n(r_n)\mathbf{R}_n^v\mathbf{L}_n(r_n)^T, \quad (12)$$

$$\mathbf{m}_{n|n-1}^i(r_n) = \mathbf{F}_n(r_n)\mathbf{m}_{n-1|n-1}^i, \quad (13)$$

$$\mathbf{P}_{n|n-1}^i(r_n) = \mathbf{F}_n(r_n)\mathbf{P}_{n-1|n-1}^i\mathbf{F}_n(r_n)^T + \mathbf{G}_n(r_n)\mathbf{Q}_n\mathbf{G}_n(r_n)^T, \quad (14)$$

so

$$p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i) = \sum_{r_n} \mathcal{N}(\mathbf{y}_n; \tilde{\mathbf{y}}_n^i(r_n); \mathbf{S}_n^i(r_n)) p(r_n | r_{n-1}^i). \quad (15)$$

In this case the new weights are given by

$$w_n^i = \frac{w_{n-1}^i p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i)}{\sum_{i=1}^N w_{n-1}^i p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i)} \quad (16)$$

and thus do not depend on particles $\{\mathbf{r}_{0:n-1}^i\}_{i=1}^N$. Finally from (5), estimator $\hat{\Theta}_n^{\text{RB}}$ deduced from this RBPF is given by

$$\hat{\Theta}_n^{\text{RB}} = \sum_{i=1}^N w_n^i \Phi(\mathbf{r}_{0:n}^i). \quad (17)$$

3. A FURTHER RB STEP

We assume as previously that $p(\mathbf{r}_{0:n-1} | \mathbf{y}_{0:n-1})$ is approximated by a set of weighted samples $\{w_{n-1}^i, \mathbf{r}_{0:n-1}^i\}_{i=1}^N$. From model (1) we have

$$p(\mathbf{x}_n, r_n | \mathbf{y}_{0:n-1}) = \sum_{\mathbf{r}_{0:n-1}} p(r_n | r_{n-1}) p(\mathbf{r}_{0:n-1} | \mathbf{y}_{0:n-1}) \times$$

$$\int f_{n|n-1}(\mathbf{x}_n | r_n, \mathbf{x}_{n-1}) p(\mathbf{x}_{n-1} | \mathbf{r}_{0:n-1}, \mathbf{y}_{0:n-1}) d\mathbf{x}_{n-1}. \quad (18)$$

If we plug the Monte Carlo approximation of $p(\mathbf{r}_{0:n-1} | \mathbf{y}_{0:n-1})$ in (18), we get an approximation $\tilde{p}(\mathbf{x}_n, r_n | \mathbf{y}_{0:n-1})$ of $p(\mathbf{x}_n, r_n | \mathbf{y}_{0:n-1})$ given by:

$$\begin{aligned} \tilde{p}(\mathbf{x}_n, r_n | \mathbf{y}_{0:n-1}) &= \sum_{i=1}^N w_{n-1}^i p(r_n | r_{n-1}^i) \times \\ &\int f_{n|n-1}(\mathbf{x}_n | r_n, \mathbf{x}_{n-1}) \underbrace{p(\mathbf{x}_{n-1} | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n-1})}_{\mathcal{N}(\mathbf{m}_{n-1|n-1}^i, \mathbf{P}_{n-1|n-1}^i)} d\mathbf{x}_{n-1} \\ &= \sum_{i=1}^N w_{n-1}^i p(r_n | r_{n-1}^i) \xi_{n|n-1}^i(\mathbf{x}_n, r_n) \end{aligned} \quad (19)$$

where

$$\xi_{n|n-1}^i(\mathbf{x}, r) = \mathcal{N}(\mathbf{x}; \mathbf{m}_{n|n-1}^i(r); \mathbf{P}_{n|n-1}^i(r)), \quad (20)$$

where $\mathbf{m}_{n|n-1}^i(r)$ and $\mathbf{P}_{n|n-1}^i(r)$ are respectively defined in (13) and (14).

Next, according to Baye's rule we have (here \mathcal{N}' stands for numerator):

$$p(\mathbf{x}_n, r_n | \mathbf{y}_{0:n}) = \frac{g_n(\mathbf{y}_n | \mathbf{x}_n, r_n) p(\mathbf{x}_n, r_n | \mathbf{y}_{0:n-1})}{\sum_{r_n} \int \mathcal{N}'(\mathbf{x}_n, r_n) d\mathbf{x}_n}. \quad (21)$$

Plugging (19) into (21) we get an approximation $\tilde{p}(\mathbf{x}_n, r_n | \mathbf{y}_{0:n})$ of $p(\mathbf{x}_n, r_n | \mathbf{y}_{0:n})$:

$$\begin{aligned} \tilde{p}(\mathbf{x}_n, r_n | \mathbf{y}_{0:n}) &= \\ &\frac{\sum_{i=1}^N w_{n-1}^i p(r_n | r_{n-1}^i) g_n(\mathbf{y}_n | \mathbf{x}_n, r_n) \xi_{n|n-1}^i(\mathbf{x}_n, r_n)}{\sum_{r_n} \int \mathcal{N}'(\mathbf{x}_n, r_n) d\mathbf{x}_n}. \end{aligned} \quad (22)$$

Since $g_n(\mathbf{y}_n | \mathbf{x}_n, r_n) = \mathcal{N}(\mathbf{y}_n; \mathbf{H}_n(r_n) \mathbf{x}_n; \mathbf{L}_n(r_n) \mathbf{R}_n^v \times \mathbf{L}_n^T(r_n))$, we have

$$\begin{aligned} \tilde{p}(\mathbf{x}_n, r_n | \mathbf{y}_{0:n}) &= \\ &\frac{\sum_{i=1}^N w_{n-1}^i p(r_n | r_{n-1}^i) \mathcal{N}(\mathbf{y}_n; \tilde{\mathbf{y}}_n^i(r_n); \mathbf{S}_n^i(r_n)) \xi_n^i(\mathbf{x}_n, r_n)}{\sum_{r_n} \sum_{i=1}^N w_{n-1}^i p(r_n | r_{n-1}^i) \mathcal{N}(\mathbf{y}_n; \tilde{\mathbf{y}}_n^i(r_n); \mathbf{S}_n^i(r_n))} \end{aligned}$$

where $\tilde{\mathbf{y}}_n^i(r_n)$ and $\mathbf{S}_n^i(r_n)$ are defined respectively in (11) and (12), and

$$\xi_n^i(\mathbf{x}_n, r_n) = \mathcal{N}(\mathbf{x}_n; \mathbf{m}_{n|n}^i(r_n); \mathbf{P}_{n|n}^i(r_n)), \quad (23)$$

$$\mathbf{m}_{n|n}^i(r_n) = \mathbf{m}_{n|n-1}^i(r_n) + \mathbf{K}_n^i(r_n) \tilde{\mathbf{y}}_n^i(r_n), \quad (24)$$

$$\mathbf{P}_{n|n}^i(r_n) = (\mathbf{I} - \mathbf{K}_n^i(r_n) \mathbf{H}_n(r_n)) \mathbf{P}_{n|n-1}^i(r_n), \quad (25)$$

$$\mathbf{K}_n^i(r_n) = \mathbf{P}_{n|n-1}^i(r_n) \mathbf{H}_n^T(r_n) \mathbf{S}_n^i(r_n)^{-1}. \quad (26)$$

Finally, setting for all $1 \leq i \leq N$

$$w_n^i(r_n) = w_{n-1}^i p(r_n | r_{n-1}^i) \mathcal{N}(\mathbf{y}_n; \tilde{\mathbf{y}}_n^i(r_n); \mathbf{S}_n^i(r_n)), \quad (27)$$

we get a new estimator $\hat{\Theta}_n^{\text{RB-2}}$ of Θ_n given by

$$\hat{\Theta}_n^{\text{RB-2}} = \frac{\sum_{r_n} \sum_{i=1}^N w_n^i(r_n) \Phi_n(\mathbf{r}_{0:n-1}^i, r_n)}{\sum_{r_n} \sum_{i=1}^N w_n^i(r_n)} \quad (28)$$

One can next derive the approximation of $p(\mathbf{r}_{0:n} | \mathbf{y}_{0:n})$ for the next iteration by using the optimal ID (10), since $p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i, r_n)$ and $p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i)$ have already been computed.

4. DISCUSSION

4.1. Properties of $\hat{\Theta}_n^{\text{RB-2}}$

Let us now compare the conditional mean and variance of $\hat{\Theta}_n^{\text{RB}}$ and $\hat{\Theta}_n^{\text{RB-2}}$. From (16), $E(\hat{\Theta}_n^{\text{RB}} | \{\mathbf{r}_{0:n-1}^i\}_{i=1}^N, \mathbf{y}_{0:n}) = \sum_{i=1}^N w_n^i \sum_{r_n} \Phi_n(\mathbf{r}_{0:n-1}^i, r_n) \times p(r_n | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})$. According to (16) and (10)

$$\begin{aligned} &w_n^i \times p(r_n | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n}) \\ &= \frac{\overbrace{w_n^i(r_n)}^{\sum_{i=1}^N w_n^i(r_n)} p(r_n | r_{n-1}^i) p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i, r_n)}{\sum_{i=1}^N \sum_{r_n} \mathcal{N}'(r_n, \mathbf{x}_n)} \\ &= \frac{w_n^i(r_n)}{\sum_{i=1}^N \sum_{r_n} w_n^i(r_n)} \end{aligned}$$

where $w_n^i(r_n)$ is defined in (27). So

$$E(\hat{\Theta}_n^{\text{RB}} | \{\mathbf{r}_{0:n}^i\}_{i=1}^N, \mathbf{y}_{0:n}) = \hat{\Theta}_n^{\text{RB-2}}. \quad (29)$$

On the other hand, let us remark that

$$\text{var}(\hat{\Theta}_n^{\text{RB-2}} | \{\mathbf{r}_{0:n-1}^i\}_{i=1}^N, \mathbf{y}_{0:n}) = 0. \quad (30)$$

Consequently, removing the dependence in $\{\mathbf{r}_{0:n-1}^i\}_{i=1}^N$ which have been sampled according to the same distribution for both estimators, we have

$$E(\hat{\Theta}_n^{\text{RB}} | \mathbf{y}_{0:n}) = E(\hat{\Theta}_n^{\text{RB-2}} | \mathbf{y}_{0:n}).$$

Applying the RB equality:

$$\text{var}(X) = \text{var}(E(X|Y)) + E(\text{var}(X|Y)),$$

to (29), we get $\text{var}(\hat{\Theta}_n^{\text{RB}} | \mathbf{y}_{0:n}) = \text{var}(\hat{\Theta}_n^{\text{RB-2}} | \mathbf{y}_{0:n}) + E(\text{var}(\hat{\Theta}_n^{\text{RB}} | \{\mathbf{r}_{0:n-1}^i\}_{i=1}^N, \mathbf{y}_{0:n}) | \mathbf{y}_{0:n})$. In conclusion our RB estimator $\hat{\Theta}_n^{\text{RB-2}}$ outperforms the classical one $\hat{\Theta}_n^{\text{RB}}$ recalled in Section 2 since both estimators have the same mean but the variance of $\hat{\Theta}_n^{\text{RB-2}}$ is lower or equal to that of $\hat{\Theta}_n^{\text{RB}}$.

4.2. Extra computational cost

The counterpart of this variance reduction is that it involves an extra computational cost. However this extra computational cost is not prohibitive, as we now see. Remember that computing $\hat{\Theta}_n^{\text{RB-2}}$ involves the computation of $\{w_n^i(r_n)\}_{i=1}^N$ and $\{\Phi_n(\mathbf{r}_{0:n-1}^i, r_n)\}_{i=1}^N$ for all different values taken by r_n .

First, observe that both estimators need the computation of (11)-(14). The only difference comes from the computation of $\{\Phi_n(\mathbf{r}_{0:n-1}^i, r_n)\}_{i=1}^N$ which has to be done for all r_n in the case of our estimator $\hat{\Theta}_n^{\text{RB-2}}$, but only for the new sampled particle r_n^i in the case of the original estimator $\hat{\Theta}_n^{\text{RB}}$. However, the computation of $\Phi_n(\mathbf{r}_{0:n-1}^i, r_n)$ is just given by the KF which has been partially computed in (11)-(14). So if one needs to compute $\Phi_n(\mathbf{r}_{0:n-1}^i, r_n)$ for all $i, 1 \leq i \leq N$ and if K is the possible number of values taken by r_n our RB solution involves the computation of $3 \times N \times (K - 1)$ extra equations, see (24), (25) and (26). However, if $\phi(\mathbf{x}_n) = \mathbf{x}_n$, only the computation of (24) and of (26) is necessary.

5. EXTENSION TO NONLINEAR JMSS

Our new RB estimator was developed for JMLS and does not need any additional assumptions as compared to the original RBPF [4]. In a nonlinear model, it is no longer possible to compute $w_n^i(r_n)$ and $\Phi_n(\mathbf{r}_{0:n-1}^i, r_n)$ for all i , $1 \leq i \leq N$. However, remember that their computation relies on the KF. So if given r_n non linearities are not too severe, it remains possible to compute our further RB estimator $\hat{\Theta}_n^{\text{RB-2}}$ using well-known approximate techniques such the Extended KF (EKF) or the Unscented KF (UKF) [12].

Now, if nonlinearities are severe one needs to adapt our RB methodology to the first class of SMC methods for JMSS, see section 1. Remember that in this case we do not compute $\Phi_n(\mathbf{r}_{0:n})$ but rather look for propagating an approximation of $p(\mathbf{x}_n, \mathbf{r}_{0:n} | \mathbf{y}_{0:n})$ [4] [13]. Let us assume that the set $\{w_{n-1}^i, \mathbf{r}_{0:n-1}^i, \mathbf{x}_{n-1}^i\}_{i=1}^N$ is a discrete approximation of $p(\mathbf{x}_{n-1}, \mathbf{r}_{0:n-1} | \mathbf{y}_{0:n-1})$. We next follow the approach of Section 3. If we inject this discrete approximation in (18) and incorporate the new observation \mathbf{y}_n (see (21)) then we get an approximation $\tilde{p}(\mathbf{x}_n | \mathbf{y}_{0:n})$ (or simply $\tilde{p}_{n|n}$) of $p(\mathbf{x}_n | \mathbf{y}_{0:n})$ given by

$$\tilde{p}_{n|n}(\mathbf{x}_n) = \frac{\sum_{r_n} \sum_{i=1}^N \bar{w}_n^i(r_n) p(\mathbf{x}_n | \mathbf{x}_{n-1}^i, \mathbf{y}_n, r_n)}{\sum_{r_n} \sum_{i=1}^N w_n^i(r_n)} \quad (31)$$

where

$$\begin{aligned} \bar{w}_n^i(r_n) &= w_{n-1}^i p(r_n | r_{n-1}^i) p(\mathbf{y}_n | \mathbf{x}_{n-1}^i, r_n), \\ p(\mathbf{y}_n | \mathbf{x}_{n-1}, r_n) &= \int g_n(\mathbf{y}_n | \mathbf{x}_n, r_n) \times \\ &\quad f_{n|n-1}(\mathbf{x}_n | \mathbf{x}_{n-1}, r_n) d\mathbf{x}_n, \\ p(\mathbf{x}_n | \mathbf{x}_{n-1}, \mathbf{y}_n, r_n) &= \frac{g_n(\mathbf{y}_n | \mathbf{x}_n, r_n) f_{n|n-1}(\mathbf{x}_n | \mathbf{y}_n, r_n)}{p(\mathbf{y}_n | \mathbf{x}_{n-1}, r_n)}. \end{aligned}$$

So an RB estimator of Θ_n is given by:

$$\hat{\Theta}_n^{\text{RB-2}} = \frac{\sum_{i=1}^N \sum_{r_n} \bar{w}_n^i(r_n) E(\phi(\mathbf{x}_n) | \mathbf{x}_{n-1}, \mathbf{y}_n, r_n)}{\sum_{i=1}^N \sum_{r_n} \bar{w}_n^i(r_n)}. \quad (32)$$

The computation of $\hat{\Theta}_n^{\text{RB-2}}$ for nonlinear JMSS requires those of $p(\mathbf{y}_n | \mathbf{x}_{n-1}, r_n)$ and $E(\phi(\mathbf{x}_n) | \mathbf{x}_{n-1}, \mathbf{y}_n, r_n)$. Let us now discuss on some approximation techniques when these terms are not computable.

- One can approximate $p(\mathbf{y}_n | \mathbf{x}_{n-1}, r_n)$ and the first order moments of $p(\mathbf{x}_n | \mathbf{x}_{n-1}, \mathbf{y}_n, r_n)$ by using local linearizations or the Unscented Transformation [12], and next compute an approximation of $\hat{\Theta}_n^{\text{RB-2}}$. In the first paragraph of this section, we proposed approximations based on these techniques. Contrary to the first paragraph where they were used to run an EKF/UKF, here they are just used to compute an approximation of the estimator, and we do not propagate this approximation in time. The advantage is that we avoid to propagate numerical errors in time.

- The product

$$A_n = p(r_n | r_{n-1}) p(\mathbf{y}_n | \mathbf{x}_{n-1}, r_{n-1}) \times E(\phi(\mathbf{x}_n) | \mathbf{x}_{n-1}, \mathbf{y}_n, r_n) \quad (33)$$

can be rewritten as

$$A_n = p(r_n | r_{n-1}) \times \int \phi(\mathbf{x}_n) f_{n|n-1}(\mathbf{x}_n | \mathbf{x}_{n-1}, r_n) g_n(\mathbf{y}_n | \mathbf{x}_n, r_n) d\mathbf{x}_n, \quad (34)$$

so if $p(\mathbf{y}_n | \mathbf{x}_{n-1}, r_{n-1}) E(\phi(\mathbf{x}_n) | \mathbf{x}_{n-1}, \mathbf{y}_n, r_n)$ is not computable, then one can sample $\mathbf{x}_n^{r_n, i} \sim f_{n|n-1}(\mathbf{x}_n | \mathbf{x}_{n-1}^i, r_n)$ for a given r_n , and approximate the RB estimator $\hat{\Theta}_n^{\text{RB-2}}$ by

$$\hat{\Theta}_n^{\text{RB-3}} = \frac{\sum_{i=1}^N \sum_{r_n} \hat{w}_n^i(r_n) \phi(\mathbf{x}_n^i)}{\sum_{i=1}^N \sum_{r_n} \hat{w}_n^i(r_n)}. \quad (35)$$

where

$$\hat{w}_n^i(r_n) = p(r_n | r_{n-1}^i) g_n(\mathbf{y}_n | \mathbf{x}_n^{r_n, i}, r_n) \quad (36)$$

which are now computable whatever the considered model. Contrary to the classical SMC methods for nonlinear JMSS where we sample $\{\mathbf{x}_n^i, r_n^i\}_{i=1}^N$ [4] [13], we have just sampled $\{\mathbf{x}_n^{r_n, i}\}_{i=1}^N$ for all r_n to compute an estimator of Θ_n .

6. SIMULATIONS

We compute the empirical mean square error (MSE) at each time step, averaged on $P = 200$ simulations, and defined by $\text{MSE}(n) = \frac{1}{P} \sum_{i=1}^P \|\hat{\Theta}_n^j - \Theta_n^j\|^2$. Here $\|\cdot\|$ is the euclidean norm, Θ_n^j is the true mean at j -th realization computed by a PF with $N = 10^5$ particles, and $\hat{\Theta}_n^j$ denotes either the estimate of Θ_n^j computed by the RBPF with optimal ID or by our algorithm.

We test our algorithm for JMLS. We track a maneuvering target described by its position and velocity in Cartesian coordinates, $\mathbf{x}_n = [p_x, \dot{p}_x, p_y, \dot{p}_y]^T$. The model is described by equations (6)-(8); mode r_n represents the behavior of the target (straight, left turn or right turn) and parameters of the model are

$$\mathbf{F}(r) = \begin{bmatrix} 1 & \frac{\sin(\omega_r T)}{\omega_r} & 0 & \frac{1 - \cos(\omega_r T)}{\omega_r} \\ 0 & \cos(\omega_r T) & 0 & -\sin(\omega_r T) \\ 0 & \frac{1 - \cos(\omega_r T)}{\omega_r} & 1 & \frac{\sin(\omega_r T)}{\omega_r} \\ 0 & \sin(\omega_r T) & 0 & \cos(\omega_r T) \end{bmatrix},$$

$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, where $\omega_r = 0 \text{ rad.s}^{-1}$ when $r = 1$, $\omega_r = 3\pi/180 \text{ rad.s}^{-1}$ when $r = 2$ and $\omega_r = -3\pi/180 \text{ rad.s}^{-1}$ when $r = 3$, $\mathbf{G}_k(r_k) = \mathbf{I}_{4 \times 4}$, $\mathbf{L}_k(r_k) = \mathbf{I}_{2 \times 2}$,

$$\mathbf{Q} = \sigma_v^2 \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} & 0 & 0 \\ \frac{T^2}{2} & T & 0 & 0 \\ 0 & 0 & \frac{T^3}{3} & \frac{T^2}{2} \\ 0 & 0 & \frac{T^2}{2} & T \end{bmatrix} \text{ and } \mathbf{R} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}.$$

We set $T = 2s$, $\sigma_v = 3m^2/sec^3$ and $\sigma_x = \sigma_y = 10m$. The Markovian transition probability is $p(r_k|r_{k-1}) = 0.4$ if $r_k = r_{k-1}$ and $p(r_k|r_{k-1}) = 0.3$ otherwise. A realization of the target tracking scenario is displayed in Fig. 1. We compute $\hat{\Theta}_n^{RB}$ and $\hat{\Theta}_n^{RB-2}$. Since both algorithms are

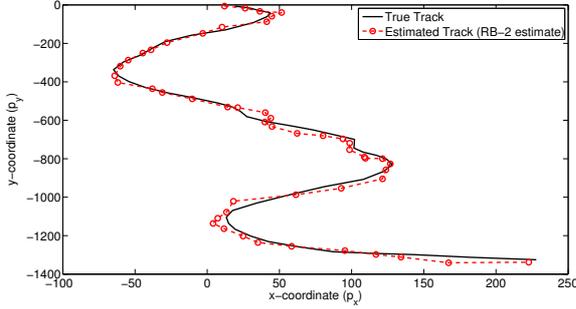


Fig. 1. Target Tracking Scenario - Jump Markov Linear Model

computationally intensive (a KF is run for each particle), we just take 100 particles and we perform a resampling step at each time step. Results are displayed in Fig. 2(a) and as expected $\hat{\Theta}_n^{RB-2}$ outperforms $\hat{\Theta}_n^{RB}$. As we see, the gap is important. This is because we use few particles and because the Markovian transition probabilities are close. Indeed, remember from (29) that

$$\begin{aligned} \hat{\Theta}_n^{RB-2} &= E(\hat{\Theta}_n^{RB} | \mathbf{r}_{0:n-1}^{1:N}, \mathbf{y}_{0:n}) \\ &= \sum_{i=1}^N w_n^i \sum_{r_n} \Phi_n(\mathbf{r}_{0:n-1}^i, r_n) p(r_n | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n}), \end{aligned}$$

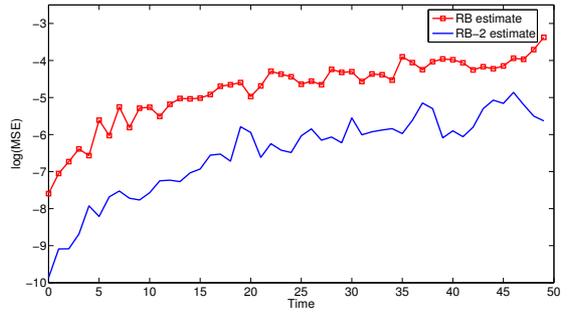
in which $p(r_n | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})$ is the optimal ID in (10). So for each i , $\hat{\Theta}_n^{RB}$ is the estimator built by sampling one particle $r_n^i \sim p(r_n | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})$, whereas $\hat{\Theta}_n^{RB-2}$ uses the mean of $\Phi_n(\mathbf{r}_{0:n-1}^i, r_n)$ according to $p(r_n | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})$. So $\hat{\Theta}_n^{RB-2}$ will really outperform $\hat{\Theta}_n^{RB}$ if the a posteriori distribution $\{p(r_n | \mathbf{r}_{0:n-1}^i, \mathbf{y}_{0:n})\}_{r_n}$ is almost uniform for all $1 \leq i \leq N$. According to (10), this can happen when $p(\mathbf{y}_n | \mathbf{y}_{0:n-1}, \mathbf{r}_{0:n-1}^i, r_n)$ and $p(r_n | r_{n-1}^i)$ do not depend too much on r_n , i.e. when the dependency of the model in MC $\{r_n\}$ is weak and the Markovian transition probabilities are close.

In paragraph 4.2, we discussed on the extra computational cost for $\hat{\Theta}_n^{RB-2}$. It is thus of interest to take into account the computation time. So we compute the efficiency for both estimators at each time step defined as [14]

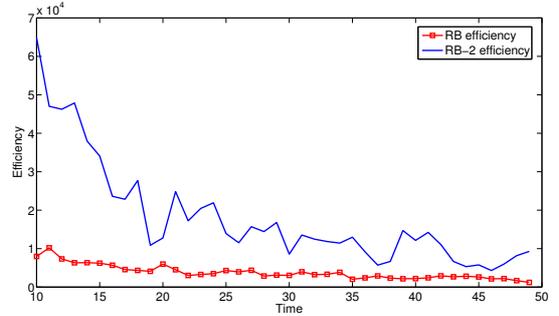
$$\text{Eff}(n) = \frac{1}{\text{MSE}(n)E(C(n))}, \quad (37)$$

where $C(n)$ is the CPU time to compute both estimators. The efficiency is displayed for $10 \leq n \leq 50$ (for smaller values of n , the efficiency of $\hat{\Theta}_n^{RB-2}$ is approximately 10^5) in Fig. 2(b) and indeed we see that $\hat{\Theta}_n^{RB-2}$ is of practical interest since its efficiency is greater than that of $\hat{\Theta}_n^{RB}$.

Let us now compare $\hat{\Theta}_n^{RB-2}$ computed with $N_2 = 50$ particles to the the classical RB estimator $\hat{\Theta}_n^{RB}$ computed



(a)



(b)

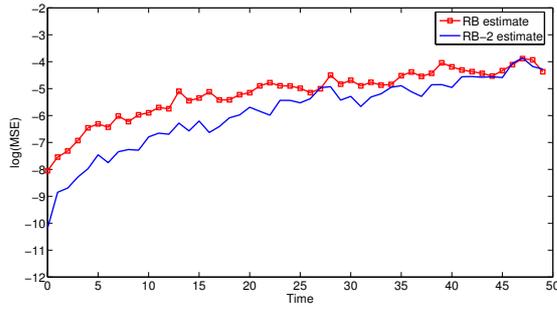
Fig. 2. MSE (a) - Efficiency (b) - JMLS - Close Markovian transition probabilities - $N = 100$ $\phi(\mathbf{x}_n) = \mathbf{x}_n$

with $N_1 = 200$ particles. MSE and efficiency are displayed in Fig. 3. With 50 particles only $\hat{\Theta}_n^{RB-2}$ still outperforms $\hat{\Theta}_n^{RB}$, even if the gap is weaker in terms of MSE. However, this is corrected by the computational cost since we used less particles to compute $\hat{\Theta}_n^{RB-2}$. As we see in Fig. 3(b), the efficiency is not responsive to the decrease of the number of particles.

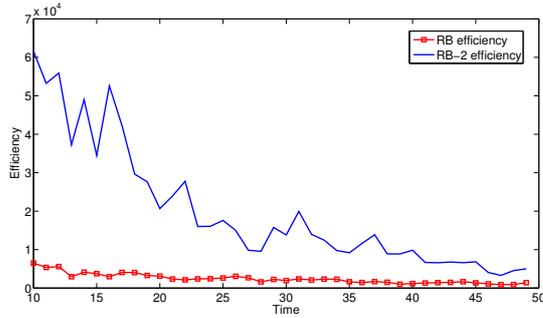
We next increase the dependency of the model in $\{r_n\}$ by setting $\omega_r = 8\pi/180 \text{ rad.s}^{-1}$ when $r = 2$ and $\omega_r = -8\pi/180 \text{ rad.s}^{-1}$ when $r = 3$, and we take more dispersed transition probabilities by setting $p(r_k|r_{k-1}) = 0.6$ if $r_k = r_{k-1}$ and $p(r_k|r_{k-1}) = 0.2$ otherwise. Results are displayed in Fig. 4 and as expected the gap between $\hat{\Theta}_n^{RB-2}$ and $\hat{\Theta}_n^{RB}$ is less important than in the previous simulation. However, $\hat{\Theta}_n^{RB-2}$ remains more efficient than $\hat{\Theta}_n^{RB}$ as we see in Fig. 4(b).

7. CONCLUSION

In this paper we derived a new RB estimator of a moment of a function of interest in JMLS. We proved that this estimator is actually an RB version of the already RB estimator with optimal ID, and as such outperforms the original estimator. We discussed on the extra computational cost needed by our approach, and we extended our algorithm for nonlinear JMSS. Our approach was finally validated by simulations.

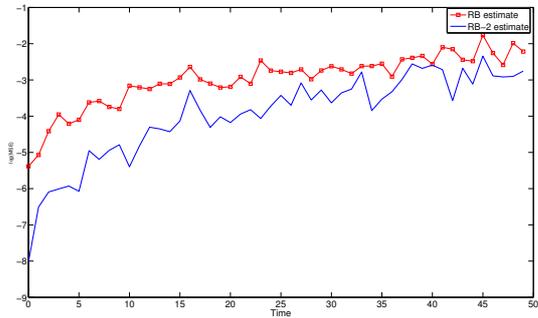


(a)

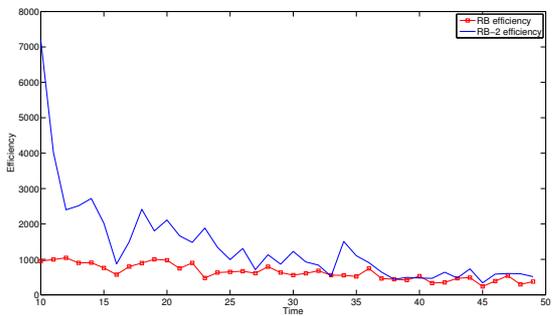


(b)

Fig. 3. MSE (a) - Efficiency (b) - JMLS - Close Markovian transition probabilities - $N_1 = 200$, $N_2 = 50$ - $\phi(\mathbf{x}_n) = \mathbf{x}_n$



(a)



(b)

Fig. 4. MSE (a) - Efficiency (b) - JMLS - Dispersed Markovian transition probabilities - $N = 100$ - $\phi(\mathbf{x}_n) = \mathbf{x}_n$

8. REFERENCES

- [1] G. Ackerson and K. Fu, "On state estimation in switching environments," *IEEE Transactions on Automatic Control*, vol. AC-15, pp. 10–17, 1970.
- [2] H. Akashi and H. Kumamoto, "Random sampling approach to state estimation in switching environments," *Automatica*, vol. 13, pp. 429–434, 1977.
- [3] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.
- [4] A. Doucet, N. J. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump Markov linear systems," *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 613–24, March 2001.
- [5] R. Chen and J. S. Liu, "Mixture Kalman filters," *J. R. Statist. Soc. B*, vol. 62, pp. 493–508, 2000.
- [6] T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear non-linear state-space models," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2279–2289, 2005.
- [7] F. Lindsten, T. Schön, and J. Olsson, "An explicit variance reduction expression for the Rao-Blackwellized particle filter," in *18th World Congress of the International Federation of Automatic Control (IFAC)*, August 2011.
- [8] J. K. Tugnait, "Adaptive estimation and identification for discrete systems with Markov jump parameters," *IEEE Transactions on Automatic Control*, vol. 27, no. 5, pp. 1054–65, October 1982.
- [9] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New-York: John Wiley and sons, 2001.
- [10] A. Kong, J. Liu, and W. H. Wong, "Sequential Imputations and Bayesian Missing Data Problems," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 278–88, March 1994.
- [11] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputation," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 567–76, June 1995.
- [12] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," in *Proceedings of the IEEE*, vol. 92, March 2004, pp. 401–422.
- [13] C. Andrieu, M. Davy, and A. Doucet, "Efficient particle filtering for jump Markov systems," *IEEE trans. on Signal Processing*, vol. 51, pp. 1762–1770, 2002.
- [14] P. L'Ecuyer, "Efficiency improvement and variance reduction," in *Winter Simulation Conference 1994*, 1994, pp. 122–132.