Report on the manuscript JSIG-D-17-00075 - Online Sequential Monte Carlo smoother for partially observed stochastic differential equations

The manuscript proposes a new sequential Monte Carlo methodology for the smoothing problem in partially observed SDEs. More specifically, the authors propose an extension of the PaRIS algorithm based on an unbiased estimator for the diffusion transition density obtained via the Poisson Estimator. The authors develop a consistent estimator for smoothed expectations of additive functionals of the diffusion. The proposed SMC algorithm, called GRand PaRIS, has a complexity that may be linear or quadratic in the number of particles, depending on the SDE being considered and on the specification of a given function. The proposed methodology is shown to outperforms the fixed lag smoother of [23] in two examples.

The manuscript is well presented and mathematically rigorous and provides significant contribution to the literature. Nevertheless, I believe there are some important issues that should be addressed in order to make it suitable for publication.

1. The authors should clearly characterise the class of processes for which the proposed methodology can be applied. In particular, three issues should be addressed when presenting the process in (1). First, condition $\alpha(x) = \nabla_x A(x)$ should be emphasised with its validity discussed in the univariate and multivariate cases. Second, it should be said that processes with general diffusion coefficients may be considered as long as a suitable transformation to lead to a unit coefficient is available. This should also be discussed in the univariate and multivariate cases. Third, it could mention that ϕ (which depends on α) should be bounded below in the state space of X. It would also be nice to give an insight as to why all these conditions are necessary.

2. When presenting equation (2), the authors should explain the importance of approximating those type of functionals.

3. General practical strategies to specify ϑ_k, p_k, $\hat{\sigma}_+^k$ in the SDE context should be discussed.

4. Conditions (A1)-(A3) should be carefully discussed. How restrictive are they? For example, what if $\phi(w_s)$ (or L_w) is not bounded below? Is there any general solution in this case?

5. The authors should discuss how to obtain L_w and U_w, and the efficiency of their algorithm in terms of these bounds. In particular, they should reference the layered BB construction - Beskos et al. (2008). Also, they can take advantage of the factorisation (into two terms) in (9) to perform the reject-accept step of the algorithm in Lemma 1 using two coins - note that the first term does not depend on BB points so a fail in the first coin would make the simulation of the BB points unnecessary.

6. It is worth mentioning that (A3) is also satisfied for the EA3 class of diffusions. See Beskos et al. (2008).

7. Concerning the paragraph starting in page 8 and ending in page 9. Can the authors provide some analytical result about when that strategy is better as a function of the rejection sampling global accep. prob. and the cost to sample (layered) Brownian bridges?

8. Item 3 above should include a discussion regarding assumptions H1 and H2.

9. Can anything be said about the bias of the proposed estimator for finite N? Is it only due to the form of Λ_k^N?

10. What can be said about a CLT for the proposed estimator?

11. It is worth mentioning that, for the log-growth model, X is a function of σ and what are the consequences of that in a context where σ is an unknown parameter.

12. I quite like the choice of $Q(\theta, \theta)$ in the examples. It would be very interesting to have an implementation of the actual EM algorithm for parameter estimation in the examples provided, using GRand Paris on the E step.

Minor comments:

1. page 3, line 36: The authors should clearly state that the proposed methodology does not rely on time discretisation approximations.

2. page 5, line 21: computing $\rightarrow$ computed.

3. page 8, line 52: I suggest "... L_w is bounded below almost surely, ..."

4. page 11, line 10: does the fixed lag algorithm have the same cost for all lag values? What is the computational time of one run of GRand PaRIS for the examples presented?

5. page 16, line 49: For the proof to be valid, the induction result needs to be established for $k \geq 0$.

References:

Beskos, A., O. Papaspiliopoulos, G. O. Roberts (2008). A new factorisation of diffusion measure and sample path reconstruction. *Methodology and Computing in Applied Probability* 10, 85-104.