Response to Referee Comments: Parallel Adaptive Importance Sampling

September 7, 2017

First of all, we would like to thank the referees for their time in reading and commenting on our manuscript. In this document we will address each of the points raised, and describe any relevant changes that have been made to the paper.

Referee 1 comments

Comment: *The PAIS algorithm does exhibits interesting ideas to combine the information obtained from an ensemble of proposals. Furthermore, it seems quite flexible since several proposals and re-sampling methods can be used.*

However, the examples somehow fail to convince that the algorithm can be used in a wide range of challenging inverse problems. The approach is limited by the dimensionality of the problem..

Answer: We thank the referee for their comments, but regarding the final comments, we can't agree. We are very open about the fact that this approach cannot be used for high dimensional problems, but there are a plethora of lower dimensional inverse problems which exhibit structure in the posterior, such as multimodality or high correlations, for which existing MCMC methods struggle to converge in reasonable time. Dimensionality is not the only challenge faced in inverse problems, and these methods have been developed to tackle problems of a different nature. In particular, this method was designed with a view to sampling from inverse problems arising in Biology, where challenging posteriors exist on relatively low dimensional state spaces. We have added some words on this to clarify further the intended type of problem that this method is intended for in Section ****.

Comment: *For mixture problems, if the mixture has to many elements (or the number of elements is unknown), then this scheme will have difficulties..*

Answer: This is true, as would RWMH, MALA, HMC and a whole range of other methods. I would be happy for the referee to indicate a method which would comfortably deal with a mixture problem with a large number of possible known/unknown components. The example was chosen to demonstrate the remarkable performance of the algorithm (with respect to standard MCMC methods) for multimodal posteriors, and we believe it has achieved that. We have added a sentence to the end of the example to discuss the point that the referee makes.

Comment: *Also, it is not clear how it will behave in real life inverse problems where the likelihood is determined by complex physical systems that involve, for example, ODE's or PDE's. If the likelihood is expensive to evaluate, the tunning of the scaling parameters for the PAIS-RW might be a concern when the number of chains is big..*

Answer: The tuning of the scaling parameters is actually very fast, often converging within a few percentage points of the optimum within the order of 10^2 iterations. We have included some plots to demonstrate this point, and we thank the referee for bringing this issue to our attention. This is quicker than one might expect for an adaptive serial algorithm. Moreover, in a previous iteration of the manuscript, we included an example where the dynamical system of concern was a continuous time Markov chain, but this was removed following previous referee comments; we would be more than happy to reintroduce this example should the referee wish us to.

Comment: *An real data example for the PAIS algorithm is much needed. It will be interesting to see how PAIS is compared not only against a naive parallelization, but also compared against a state of the art MCMC algorithm or/and a efficient ad hoc MCMC algorithm designed for such example..*

Answer: This is the problematic one I think. Very unhelpful. There are untold methodological papers without “real life data”, such a problem requires building up a relationship with a non-mathematician, getting data, cleaning it up. It's a LOT of work.

Comment: *There aren't details about the architecture used to run the M chains in parallel. This is relevant since there are several options such as: multi-core and multi-processor computers, GPU cards, clusters, cloud computing, etc. Comments about the advantages and disadvantages of the architecture are also welcome. The same goes for details regarding the programming language used..*

Answer: We thank the referee for this comment, as we should have included these details in the manuscript. The method was implemented in C++, serially, and was run on a single core of a *****server spec here*****. We are not computer scientists, and as such, we are not in a position to comment on specific architectures. If this algorithm was to be spread over several cores, with each core computing the weight of one/more ensemble members, then issues of communication between cores would need to be addressed. Communication is only required at the point of resampling, where new positions of weights are communicated to a single core that computes the resampling, and the new ensemble positions are then sent back out. Some comments on this have been added to Section ****. However, following comments by the second referee, we have decided to reduce the emphasis on the parallelisation of the algorithm, not least because we have not implemented it in parallel ourselves. Please see our comments to the second referee for further details.

Comment: *CPU time must be reported to compare the gain of each algorithm in order to have a fair comparison. Authors claim that there is a significant gain in terms of the number of iterations needed to achieve a desire tolerance; examples in sections 7.2 and 7.3 show that PAIS saves up to 90% of the iterations. However, if the CPU time needed to run 10 PAIS iterations is ten times the time needed to run 100 RWMH iterations, then the gain is not as good as presented. PAIS iterations are more time consuming than RWMH iterations due to the re-sampling step. The authors mentioned that the re-sampling step is computationally expensive and that is why they proposed the AMR re-sampling method. CPU time is of particular importance in example 7.4 since the number of parallel chains grows from 50 to 500..*

Answer: The referee may have missed Figure **** in which we show the computational cost, in time, of the resamplers that we use in the paper, both the optimal transport resampler, and the AMR. As mentioned before in response to another of the referees comments, we had neglected to share the specs of the machine that these times were recorded on, and this has now been added. There is an additional cost in the algorithm over and above the resampler, which is the computation of the denominator in the weights, as pointed out by the second referee. If we have M members of our ensemble, then the cost of this denominator across the whole ensemble is M^2, but it can be easily parallelised. We include some timings of the cost of this aspect of the algorithm in Section ****, and a discussion of the overall additional cost of the approach. Coupled with the timings presented for the resampler, we believe that this now gives the reader a clear idea of the additional overhead costs which are incurred when using our method as opposed to a standard MCMC method. We thank the referee for their comments on this.

Referee 2 comments

Comment: *This work in an interesting contribution in the field of Monte Carlo algorithms. The paper is technically sound and well-written. Moreover, it is possible to note the effort nd the care devoted by the authors..*

Answer: We thank the referee for their kind comments.

Comment: *However, I have some concerns.*

1) The starting-point algorithm in Table 1, is very similar to the basic scheme suggested in, for instance,

V. Elvira, L. Martino, D. Luengo, M. F. Bugallo, Improving Population Monte Carlo: Alternative Weighting and Resampling Schemes, Signal Processing Vol. 131, pp. 77-91, 2017.

In practice, you use a deterministic approach for the IS weights. I believe that the authors have analyzed and studied this idea independently, but this similarity should be discussed in the introduction, at least. You can also point out that you also propose some different resampling or adaptation schemes that are not contained in the previous work..

Answer: We agree that there are similarities to this other work, and we thank the referee for their comment. Additional references have been added, along with more detailed discussions about the novel aspects of this manuscript.

Comment: *2) I do not understand why you use the word "parallel" in the name of the algorithm. In Algorithm 1, it is difficult to recognize the parallelization. The denominator in the weight that you employ, in fact, is a bottleneck for a possible parallelization. Please clarify this point (even in the text)..*

Answer: We agree with this comment from the referee, as it may not have been made clear enough. With an ensemble of states, it is possible to distribute the computation of the weights across a set of cores. We have added a comment on this to Section ****. However, as discussed in the response to the first referee, we ourselves implemented the algorithm serially. Therefore, we have decided to change the name of the algorithm to reflect this, to "Optimal Transport Adaptive Importance Sampling" or OTAIS. Changes throughout the manuscript have been made to reflect this.

Comment: *3) Please, add a table of acronyms that can help the reader. Now it is a bit confusing..*

Answer: We have added a glossary of acronyms to an appendix.

Comment: *4) The references should be completed with other important contributions (considering also the same kind of weights, in some cases):*

- O.Cappe, R.Douc, A.Guillin, J.M.Marin, C.P.Robert, Adaptive importance sampling in general mixture classes, Statistics and Computing 18 (2008) 447-459. - R. Douc, A. Guillin, J. M. Marin, C. P. Robert, Convergence of adaptive mixtures of importance sampling schemes, Annals of Statistics 35 (2007) 420-448. - R. Douc, A. Guillin, J. M. Marin, C. P. Robert, Minimum variance importance sampling via population Monte Carlo, ESAIM: Probabil-

ity and Statistics 11 (2007) 427-447. - L. Martino, V. Elvira, D. Luengo, J. Corander, Layered Adaptive Importance Sampling, Statistics and Computing, Vol. 27 (3), pp. 599-623, 2017..

Answer: We thank the referee for thier suggestions for other highly relevant references.