REVIEW OF 'VARIANCE REDUCTION FOR MCMC METHODS VIA MARTINGALE REPRESENTATIONS'

BELOMESTNY ET AL.

In this paper the authors propose a novel method for reducing the variance of MCMC ergodic averages by means of control variates. The construction of the control variates is based on a theoretical result introduced in the paper concerning a martingale representation of the Markov chain observations in terms of orthogonal basis functions with respect to the randomness inserted into the Markov chain. In principle this idea seems interesting and may lead to improvements in current methodology.

Major issues. However in my opinion there are a few issues with this manuscript. I will briefly discuss these here.

The authors make an attempt to quantify the variance reduction obtained by the proposed approach, as well as the computational complexity to achieve this variance reduction. On these topics several results are presented as Theorems and Propositions. In my opinion, these results have no immediate value for the reader without spending significant time on their interpretation. This is because these results involve highly complicated expressions which would require a serious effort to parse in the first place (this could be considered a relatively minor problem), but more importantly, depend upon various unknown quantities (e.g. conditional expectations which are not available in closed form) and are furthermore only available up to unknown multiplicative constants. This applies to Theorem 4, Theorem 5, Proposition 6 and also the results of Section 4.1 which are not phrased as a theorem. In my opinion these results can only be phrased as lemmas or perhaps as propositions. At the end of Section 6 the authors come closer to what I think is a more useful result (stating a cost of order $\varepsilon^{-1/(1-\alpha)} \log^{1/(1-\alpha)}(\varepsilon)$) and I would recommend the authors to phrase this results with clear conditions in such a way that it becomes directly accessible to the reader. Something of the form: "Suppose f satisfies... and π satisfies... Consider the ULA algorithm with stepsize ... Then there are constants ... so that the variance of the control variates estimator with parameters satisfying ... may be bounded by ....", where the bound is phrased in an easily readible way such as presented at the end of Section 6 involving constants that are defined in the statement of the theorem and which should not depend on infinite series and conditional expectations.

Furthermore, the authors present simulation studies establishing the obtained variance reduction. It is good to see that the technique has beneficial effects on the variance, but it does not become clear what was the computational effort required to achieve this benefit. The reader is left wondering whether or not the same variance reduction could have been achieved simply by drawing more MCMC samples. Also in the comparison with the other variance reduction algorithms which are considered it is not clear what are the relative computational costs of these algorithms. Furthermore I think it would be good if the simulation studies confirm the scaling results that are discussed in the theoretical sections.

To summarize these points, I think ideas presented in this manuscript have some potential to be useful, but the current formulation of both theory and experiments is unfortunately not very informative.

Other remarks.

- p. 5, Section 3: it is only very briefly discussed what would be the type of functions ϕ_j for the examples discussed earlier in the text. It would be interesting to be more specific here in the specification. For ULA (Example 1) this is done in Section 4, at the least the authors could point forward to this section. But it would also be useful to have the formulas for MALA. What would be done in the case of Example 3? Is ULA not a special case of Example 3?
- p. 10, Remark 1: here the setting $\gamma_k = \mathrm{const} k^{-\alpha}$, $\alpha \in (0,1)$ is considered, and presented as 'the most interesting case'. For many practitioners, $\alpha = 0$ would be rather natural. What can be said in this important case? This is also applies to Section 6.
- p. 11, Theorem 5: Here the functions g are assumed uniformly bounded, whereas earlier (end of Section 5) elements in Ψ were proposed to be polynomials. In particular the bound B would be equal to infinity. How can this be reconciled?

Minor suggestions and typos. Finally I will discuss several minor suggestions, issues and typos.

- p. 2, second paragraph: after the introduction of the function U it would be useful to briefly state how this function is used. In other words just put a formula like the one below to make clear to the reader the effect of subtracting the function U from f, just to be explicit and clear.

$$\hat{\pi}_n(f) = \frac{1}{n}\sum_{i=1}^{n}[f(X_i) - U(X_i)].$$

- p. 4, Example 2: purely aesthetical but I would suggest to use the LaTeX-command `\textrm` or `\text` for the phrase 'with probability'
- p. 5, final line: "generated by generated by".
- p. 10, l. -6: "Algorithm stars with..." should be "The algorithm starts with...". It happens at several places that articles ('a', 'the') are missing; please check throughout the manuscript.
- p. 11, l. 9: "Conditioning ... trajectories." is not a sentence. Use the phrase "... is based upon ..."?
- p. 14: the bound for $\pi^N_{K,n}(f)$ involves $\log(\varepsilon)$. Better would be $\log(1/\varepsilon)$ to avoid an unwanted minus sign.