

# Response to the Report by Referee 1

Denis Belomestny, Eric Moulines, and Sergei Samsonov

We thank the referee for carefully reading our manuscript and for pertinent comments which helped us to improve the paper. In the new version of our paper, we have taken all comments and suggestion into account. Below are some additional responses to the comments by Referee 1

## Main remarks

- ... *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 ... I would recommend the authors to phrase this results with clear conditions in such a way that it becomes directly accessible to the reader ... where the bound is phrased in an easily readable way involving constants that are defined in the statement of the theorem and which should not depend on infinite series and conditional expectations.*

Let us first stress that current literature on MCMC mainly focuses on convergence of MCMC algorithms in Total Variation and Wasserstein distances thus ignoring a sample error of these algorithms. Only few references studied MSE error of the Langevin algorithm (see e.g. Durmus and Moulines (2017)). In this work we propose a novel variance reduction approach which is able to reduce the statistical error of MCMC algorithms. The main advantage of our method is provable increase of efficiency for Langevin-based MCMC algorithms as measured by the cost-to-variance (product of cost and variance) criterion. Our approach is based on a new discrete time martingale representation for Markov chains which is of its own interest. Therefore we first present this basic result and then discuss in Section 4 variance reduction method based on the derived martingale representation. In Section 4 we also explain in an accessible way why our approach leads to smaller cost-to-variance characteristics. These results are based on a thorough analysis of the variance reduced estimate postponed to Section 5. In the revised version we significantly weaken assumptions and prove our main result (Theorem 10 and Corollary 11) for ULA under smoothness and convexity of the underlying potential outside a ball (assumption (H2)). Note that all assumptions needed are now

collected in Theorem 10 and Corollary 11. For convenience of the reader we also sketch the main steps of the proof right after Theorem 10. We decided to present our main theoretical result in the form of bounds for the functions  $A_{q,k}$  as these functions play a crucial role for deriving the variance bound (32).

- ... *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 ...*

In the revised version (Section 6.1) we measure the gain of our VR approach by comparing its cost-to-variance characteristic (product of the obtained variance and its cost) to one of the plain MCMC method. This characteristic quantifies the resulting variance reduction effect relative to computational effort spent.

- ... *here the setting  $\gamma_k = ck^{-\alpha}, \alpha \in (0, 1)$  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 ...*

In the revised version we decided to switch to a constant time step and present all our bounds for this case. On the one hand, this allows us to formulate our main results in a simpler form depending only on a single parameter  $\gamma$ . On the other hand, this is indeed the most relevant choice for practitioners.

- ... *here the functions  $g$  are assumed uniformly bounded, whereas earlier (end of Section 5) elements in  $\Psi$  were proposed to be polynomials. In particular the bound  $B$  would be equal to infinity. How can this be reconciled? ...*

In the revised version we skip this result because of size restrictions.