**Referee report on the Bernoulli submission BEJ1907-036RA0**

**Summary of the article** The authors study a novel variance reduction technique for MCMC methods. The aim is to numerically compute expectations of the form

$$\pi(f) = \int_{\mathcal{X}} f(x)\pi(dx)$$

where the state space $\mathcal{X}$ is possible high-dimensional and the density of $\pi(\cdot)$ is only known up to a normalization constant. In these situations, $\pi(f)$ is commonly approximated by the empirical weighted average $\pi_n(f)$ over a path of an MCMC chain. The question of constructing 'control variates' satisfying $\mathbb{E}[\xi] = 0$ and $\mathbb{E}[\xi^2] < \infty$ such that $\pi_n(f) - \xi$ has small variance, is a natural and important question.

The proposed method is based on a general martingale representation

$$f(X_p) = \mathbb{E}[f(X_p)|G_j] + \sum_{k\geq 1}\sum_{l=j+1}^{p} a_{p,l,k}(X_{l-1})\phi_k(\zeta_l),$$

where we have used the same notation as in the paper. The proposed control variate $M^N_{K,n}(f)$ is a truncated and reweighted version of the double sum (which is martingale in the index $p$) in the above display, where the weights are the natural ones – the same as in the weighted average $\pi_n(f)$.

The general representation is introduced in Section 3. In Section 4, the variance is analysed when the representation is applied to ULA, which is given by the Euler-Maruyama discretization of the diffusion

$$dY_t = -\mu(Y_t)dt + dW_t, \quad \mu = \frac{1}{2}\nabla \log d\pi.$$

A bias-variance decomposition is used. The control variate $M^N_{K,n}(f)$ contains unknown quantities which need to be estimated using Monte Carlo simulation of paths of the ULA algorithm – this is described in Section 5. In Section 6, the authors analyse the computational cost of the full numerical procedure. The main conclusion seems to be decribed on p.14: To achieve precision $\varepsilon^2$ (meaning the variance of the procedure being smaller than $\varepsilon^2$, not the bias), the computational cost is improved from order $\varepsilon^{-2/(1-\alpha)}$ to $\varepsilon^{-1/(1-\alpha)}\log^{1/(1-\alpha)}(\varepsilon)$. This can be significant in practical applications. Numerical results are provided in Section 7, which show that the in a number of simple and low-dimensional examples, the proposed method achieves significant improvement, and the performance seems to be very similar to that of the zero variance MCMC method proposed in 2010 in reference [20]of the submission.

**General comments and concerns** 1. Structure and style. The paper is carefully written – there are barely any typos. The paper also is reasonably well structured, even though it is not clear without close reading where the main results are stated. For example, one of the main results is the last part of Section 6 on p.14. If this were clearer, readability would greatly improve.

2. Mathematical content. The mathematical analysis in the paper seems interesting and novel, but still the story seems slightly unconclusive and unsatisfactory, as the needed assumptions are very strong – see the next paragraph.

3. Restrictive assumptions for the theoretical analyis. The main theorems of the paper, which seem to be Theorem 4 and the results of section 6.1, have the restrictive assumption that the drift vector field $\mu : \mathbb{R}^D \to \mathbb{R}^D$ needs to satisfy the strong convexity assumption

$$J_\mu(x) \geq cId, \quad \text{some } c > 0.$$

In typically interesting cases, such as Bayesian inverse problems and machine learning, this assumption is not satisfied. Unfortunately, even in the numerical examples provided in Section 7, this assumption is not fulfilled – it already doesn't hold for a two-component Gaussian mixture in one dimension. This is a serious restriction. In the MCMC contraction rate literature, recently a number of assumptions which relax the notion of log-concavity have been proposed – these include more general, weaker drift conditions on the vector field $\mu$, or convexity outside of a ball, or directly by assuming a log-Sobolev inequality for the target measure, see e.g. [1, 2, 3].

4. A comment on the last paragraph of Section 6. Yes, in complicated models such as PDE-related Bayesian inverse problems or machine learning, the computation of $\nabla \log(\pi)$ is expensive – but these examples are typically ones where $\log(\pi)$ does not satisfy the needed concavity assumptions.

**Conclusion** The paper is well-written and the theoretical analysis is mathematically interesting, but the assumptions of the main results are restrictive so that in even easy examples like Gaussian mixtures, they are not fulfilled anymore. I hence suggest either major revision or rejection with resubmission – if the theoretical analysis could be extended to the more general, non log-concave case, then the results would certainly gain quite some significance.

**Typos/Grammatical mistakes** There were barely any typos, but some minor grammatical mistakes

- p.2 last paragraph: other types of algorithms
- p.4 bottom of Example 2: issue with typesetting.
- p.5 second line: The definition of $D^2g$ should be given, even though it's clear it's the Hessian.
- p.7 before Prop.2: From a numerical point of view.

- p.9 second sentence of Section 4.2: we introduce a lemma
- p.10 Remark 1: Theorem 4 (remove parentheses)
- p.10 Section 5: This amounts to solving

## References

E15 [1] Eberle, A.(2015) Reflection couplings and contraction rates for diffusions. Probability Theory and Related Fields, Vol. 166, 851-866.

M17 [2] Ma, Y.-A., Chen. Y, Flammarion, N. and Jordan, M. (2018) Sampling Can be Faster than Optimization. arXiv preprint.

V19 [3] Vempala, S. and Wibisono, A. (2019) Rapid Convergence of the Unadjusted Langevin Algorithm: Isoperimetry Suffices. arXiv preprint.