

Cover Letter for
"MISSO: Minimization by Incremental Stochastic Surrogate
Optimization for Large Scale Nonconvex and Nonsmooth Problems"
Submitted to **ICLR 2021** (Grades: 3/7/7/5/6)

Our paper called "MISSO: Minimization by Incremental Stochastic Surrogate Optimization for Large Scale Nonconvex and Nonsmooth Problems" has been submitted to ICLR 2021 (submission date: October 2nd 2020 and Decisions date: January 14th 2021).

According to the full reviews received at the conference, and attached to the cover letter, we detail the following reasons for rejection along with our changes to alleviate those limitations:

- **Wall Clock Comparison:** One issue regarding the fair comparison between our method MISSO and the baseline methods such as BBB, MC-ADAM and more, is the computational time per iteration for each of those methods.

While our original submission only presented loss curves against epochs elapsed, in our experiments, we found that the tested methods involve a similar number of gradient computations per iteration (since reported every epoch), as such the wall clock time per iteration are comparable. To further support this comment, we now include additional comparison of the convergence against the running time in the main body of the paper and deferred the plot against the epochs in the supplementary material for complete and fair comparison.

- **Iteration and Sample Complexity:** While our paper provides a finite time convergence analysis of our newly introduced scheme called MISSO, see Theorem 1 for the *iteration complexity* of $\mathcal{O}(nL/\epsilon)$, several reviewers argued that in our stochastic method, involving the sampling of batches of latent variables at each iteration, it is important to quantify its *sample complexity*, meaning the number of Monte Carlo samples required at each iteration to guarantee the proven $\mathcal{O}(nL/\epsilon)$ convergence rate.

We now include in our paper a discussion on the iteration and sample complexity of our method at the end of the Theoretical section.

- **Span of Applications:** Arguably, the utility of our method was questioned in the reviewing phase. We defend that the main contribution of this paper is to propose and analyze a unifying framework for a large class of optimization algorithms which includes many well-known but not so well-studied algorithms such as *Variational Inference* and *Monte Carlo EM*. Many concrete applications leverage the power of those cited algorithms such as pharmacology or economy for the EM-type of methods, or any kind of latent variable models (Bayesian Neural Networks for classification) for the Variational Inference type of methods. We have added, upon request, some references to emphasize on the practical benefits of all the methods that our scheme MISSO encompass, such as the factorization of huge matrices (see Mensch et. al., 2017 reference).