1 We would like to thank four reviewers for their feedback. Upon acceptance, we will include in the final version (a) *a*
2 *clearer presentation of the numerical results* and (b) *missing references*. We first discuss a common concern shared by
3 **reviewer 1**, **reviewer 2**, **reviewer 4**.

4 ●●● **Novelty of The Contribution**: We want to stress on the generality of our incremental framework, which tackles a
5 *constrained*, *non-convex* and *non-smooth* optimization problem. The main contribution of this paper is to propose a
6 *unifying* framework for the analysis of a large class of optimization algorithms which indeed includes well-known but
7 not so well-studied algorithms. The major goal here is to relax the class of surrogate functions used in MISO [Mairal,
8 2015] and replace that by the respective Monte-Carlo approximations. We provide a general algorithm and global
9 convergence analysis under mild assumptions on the model and show that two examples, MLE for general latent data
10 models and Variational Inference, are its special instances.

11 Working at the crossroads of *Optimization* and *Sampling* constitues what we believe to be the novelty and the technicality
12 of our theoretical results.

13 **Reviewer 1:** We thank the reviewer for valuable comments and references. We would like to make the following
14 clarification regarding the difference with MISO:

15 **Originality:** The main contribution of the paper is to extend the MISO algorithm when the surrogate fun ctions are not
16 tractable. We motivate the need for dealing with intractable surrogate functions when nonconvex latent data models
17 are being trained. In this case, the latent structure yields an expected surrogate functions and the nonconvexity yields
18 an intractable expectation to compute. The only option is to build a stochastic surrogate function based on a MC
19 approximation.

20 **Reviewer 2:** We thank the reviewer for the useful comments. Our point-to-point response is as follows:

21 **Numerical Plots:** Due to space constraints, we only presented several dimension for the logistic parameter and the
22 mean of the latent variable. As the reviewer mentioned, we also learn the variance of those latent variables and the
23 convergence plots of those variances will be added to the rebuttal version. We made sure for all experiments that the
24 estimated parameters for each method converge to the same value. This was our only reference to claim that a method
25 is faster than the other. Then, the problem being (highly) nonconvex, indeed the estimations can get trapped in various
26 local minima. Regardless of generalization properties of the output vector of estimated parameters, our focus through
27 those numerical examples was to highlight faster convergence, in iteration, of our method.

28 **Wallclock Time**: Wallclock time per iteration is comparable for each method. Indeed the methods always only involve
29 first order computation. Yet, we acknowledge that MISSO can present some memory bottlenecks since it requires to
30 store n gradients through the run. This has not been a problem for the presented numerical examples.

31 **Parameter Tuning:** The baseline methods were tuned and presented to the best of their performances both with regards
32 to their stepsize (grid search) and minibatch size. We believe your remark refers to the first numerical example (logistic
33 regression with missing values): Regarding the stepsize, as MCEM does not have one, we indeed tuned the stepsize of
34 SAEM. Rather than c/k, common practice is to tune a parameter α such that $\gamma_k = 1/\gamma^\alpha$. We report results for SAEM
35 with the best α ($\alpha = 0.6$). Regarding batch size, for SAEM and MCEM both are full batch methods and the idea here is
36 to compare different values of minibatch size for the MISSO method to see its influence on the performances.

37 **Reviewer 3:** We thank the reviewer for valuable comments and references. We clarify the following point:

38 **Verification of the Assumptions:** TODO

39 **Reviewer 4:** We thank the reviewer for valuable comments and the numerous related references. Our point-to-point
40 response is as follows:

41 **Comparison to [Murray+, 2012] and [Tran+, 2017]:** We agree that the literature of the Bayesian Inference is rich
42 and could be more explicit in our introduction. Though, our focus is not on a method for posterior sampling purposes
43 but rather on the interlink between nonconvex optimization and sampling techniques. While our analysis cover direct,
44 and MCMC, sampling, we agree that future research directions analyzing complex bayesian inference techniques and
45 nonconvex optimization can interesting.

46 **Comparison to [Kang+, 2015]:** [Kang+, 2015] is indeed a relevant reference and will be added to the introduction
47 along the MISO [Marial, 2015] reference. Nevertheless, [Kang+, 2015] solely focus on MM scheme when the surrogate
48 functions are deterministic, *i.e.* can be computed exactly. While the objective function is nonconvex as in our work,
49 the construction of their surrogate functions does not imply any latent structure, inherent to the problem, and thus are
50 easily computed and characterized for convergence purposes. Also, their analysis requires *strong convexity* of the gap
51 between the convex surrogate and the nonconvex objective function while our analysis only requires a *smoothness-like*
52 assumption, see **H2**.