We thank all the reviewers for their time and valuable feedback. We will improve the draft based on your comments, and we hope you could raise your evaluation if we address your concerns.

Reviewer 2: (On benchmarks) We agree that including these baseline can be useful for the potential readers and we will add them in next version. We also want to notify the reviewers that it has been well tested in [2] that SVGD outperforms PBP so our baselines are actually stronger than PBP.

(On hyper-parameter tuning) We didn't tune the kernel bandwidth but simply apply the widely used median trick [2]. The thinning factor is common in MCMC and has been used for decades. We will add sensitive analysis on that. Our ablation study on the repulsive term weighting (α) is included in Appendix B.2 (we will give pointer on that in next version). Also, since that we propose a specific way to tune the α (as shown in appendix A) and its usefulness has been demonstrated as in all the experiment we use this criterion to select α, we believe the sensitive analysis of α is sufficient and the proposed criterion for selecting α is reliable in practice.

(On the Presentation) Sorry for the appendix indexing issues. Appendix A.5 is Appendix C.1. $\mathcal{B}_t$ denotes the Brownian motion and we will add definition. Our section 5 demonstrates that our Stein repulsive gradient can be applied to general dynamics as the limiting system is able to produce correct targeted distribution. As the later discretization and large particle approximation is almost identical to that in SRLD, we omit the details. We will give more discussion on that. The reference in L.100 should be referred to equation between L.80 and L.81 and we will fix it as well as the other clarity issue you mentioned.

Reviewer 3: (1, 2) We will add a related work section that collect the literature review in the current version, as well as a conclusion section in the next version.

(3, 4) Sorry for the index problem. Appendix A.4 should be Appendix C and Appendix A.5 should be Appendix C.1. We will give contextual bandit an individual section in the next version.

(5, 6) Notice that our repulsive term can be applied to any SG-MCMC and thus we believe the current experiment using the simple Langevin dynamics is able to show the usefulness of the repulsive term. We will test some more advanced SG-MCMC method.

(7) We have a synthetic experiment on sampling multi-mode distribution (see Appendix B.1 and B.3).

(8) We agree that the name repulsive is a little bit not exact, but given that this term provides repulsive force in practice, we believe this name is intuitive.

(9) Our method allows us to adjust the α to increase the magnitude of repulsive gradient compared with Langevin term. Please see line 440-443 for more details. The issue of kernel method in high dimension is a problem independent with this work and existing technique to mitigate this issue (e.g. [1]) can also be applied to our method.

(10) The density evolution is not derived by the direct application of standard Fokker-Planck equation. The derivation is standard and mostly calculation (similar to the one in Appendix A.3 of [5]) and thus we omit it previously. We will give details on the derivation in the next version.

(11, 12) We don't need that strong condition and the current statement is correct. And yes, it is θ_t in equ (4), sorry for the typo.

Reviewer 4: (W1) Thanks for the point! We agree it would be more complicate to conduct experiment on different advanced MCMC and we will add them in the next version. Due to the time limit, we are unable to give result on that currently. However, we also believe the current experiment design is sufficient and good. We believe the main focus on the experiment is to demonstrate the usefulness of the proposed Stein repulsive gradient and thus it is the best choice to work on the simple Langevin dynamics, which has the least hyper-parameters to tune.

(W2) We agree that the experiment on UCI has large variance and thus we conduct statistical testing (based on matched pair t-test, see, e.g. caption of Table 1) for all the results we report. As most result are statistically significant, we believe our result on UCI dataset is convincing. Besides, as we use different setting and data split for UCI dataset, the results in other papers are not directly comparable to ours.

We agree that it might be interesting to test the behavior of our method on large BNN model. However, due to the computation resource constraint, we are not able to maintain, e.g., 20 checkpoints of large neural nets at the same time and thus we currently are not able to do such experiment. Beside, we think the main focus of [3] and ours is different, which makes their experiment design not suitable to us. [3] aims to use Bayesian-like ensemble to improve the prediction accuracy rather than approximate the posterior (thus they only use very small number of particles that cannot approximate the posterior). Our aim is still on a better MCMC algorithm that has better posterior sample quality.

[1] Understanding and accelerating particle-based variational inference.
[2] Stein variational gradient descent: A general purpose bayesian inference algorithm.
[3] Cyclical Stochastic Gradient MCMC for Bayesian Deep Learning.
[4] Dropout as a bayesian approximation: Representing model uncertainty in deep learning.
[5] Stein Variational Gradient Descent as Gradient Flow.