We are grateful to all the reviewers for their careful and constructive assessment of our work, and in particular thank for their helpful suggestions for its improvement. We appreciate the reviewers' recognition of our key contribution in addressing the longstanding challenge of the curse of dimensionality for high-dimensional Bayesian inference problems, as well as its high relevance and significant impact in NeurIPS community.

Below are our responses to the constructive criticisms, insightful questions, and helpful suggestions of each reviewer.

To Reviewer 1: We thank for your understanding of the key property of the low-dimensionality of the subspace in which the posterior differs from the prior, and our motivation in exploiting this property by adaptively projecting the parameters into carefully constructed subspaces to use SVGD. **We also appreciate your frank expression of not understanding the fundamental limitation of SVGD and the advantages of our pSVGD over SVGD for high-dimensional inference, which led to your different reject score in contrast to all the other accept scores**. SVGD is limited because the kernel given in (9) leads to severe degeneracy of the repulsive force of SVGD in high dimensions, which makes samples collapsing to the modes of the posterior, as observed in [31, 34]. This is demonstrated by the significantly inaccurate posterior sample variances with increasing dimensions in Figure 3 for nonlinear problems (compared to DILI MCMC reference) and in Figure 5 for linear problems (compared to the ground truth) in Appendix B. In contrast, the posterior sample variance of pSVGD samples is preserved to be much more accurate than SVGD with increasing nominal dimensions as shown in both Figure 3 and 5, because pSVGD extracts the essential information, i.e., by projecting parameters to the intrinsically low-dimensional subspaces regardless of the nominal dimensions. This is also demonstrated by the much smaller testing errors of pSVGD compared to SVGD in Figure 1, 2, and 7. The advantage is also (but less) about the number of iterations or computational complexity, even though pSVGD converges faster with less cost (because of optimization in low dimensions) than SVGD as demonstrated in Figure 2. Admittedly, pSVGD is more involved to implement than SVGD. By definitions of H and Γ, the generalized eigenvalues always exist but they may not converge very rapidly for some problems. The intuition in line 100 is in fact formalized in the estimate (15). The quantity $\nabla \log \pi(w)$ in Algorithm 1 is defined in (28) and (31). We will address other minor questions. We hope that our clarifications can help you for better understanding and reevaluating our contribution.

To Reviewer 2: We appreciate your evaluation of the **novelty, high significance, and very sound and careful claims** of our paper, and especially for **your willingness to upgrade the score to full accept after improving the clarity** in Section 3 with the following revision based on your helpful suggestions. Specifically, we will add 1-2 sentences of the intuition for the gradient information matrix for H (the often used gradient-based parameter sensitivity), add 1-2 sentences of the implication by Theorem 1 (for computing the gradient of the projected posterior), and simplify the notations (by slight abuse of notation). We actually have the comparison of cost, see in the caption of Figure 2. A strong motivation for our work is that pSVN uses Hessian matrix which is often not available or too expensive to compute for complex problems. Our examples in Section 4.3 and appendix B are actually the same as those in pSVN, which can be directly compared by examining the two papers on their accuracy and convergence. We will add a short discussion on the comparison of their computational cost in the supplementary material. For more discussion on this see below in To Reviewer 3. Thank you for appreciating the COVID example. We will add a short discussion of it in the main body.

To Reviewer 3: Many thanks for your overall very comprehensive and encouraging evaluation of our paper as **sufficiently significant to warrant publication** for our methodological contribution with both theoretical and empirical justifications. On your question about the comparison of pSVGD and pSVN, please see our response above in To Reviewer 2. Moreover, we emphasize that the gradient information matrix in pSVGD only requires the gradients of the log-posterior at the pushed samples, which are already computed in SVGD, see (32). This is often the dominant cost for complex models, e.g., all the examples in this work except for the logistic regression. On the other hand, computing the Hessian in pSVN is much more demanding — (1) evaluating the full Hessian is d times more expensive than evaluating the gradient for d-dimensional parameters, (2) evaluating a Gaussian–Newton approximation of the Hessian requires Jacobian, which is s times more expensive for s-dimensional data, (3) a low-rank approximation with rank r of the Hessian is $O(r)$ times more expensive by efficient randomized algorithm. We remark that the intrinsic low-dimensional subspace is demonstrated in pSVN for a Bayesian deep learning test problem, which suggests pSVGD is also suitable for such problems, besides our five testing problems from various fields. We will address other minor issues. Thanks again for your very careful reading and **willingness to change the score** given these clarifications.

To Reviewer 4: We appreciate your evaluation that our work is **very relevant for Bayesian inference in high dimensions** and our ideas are **novel and give a new approximate inference algorithm**, and that you find the **technical presentation clear and easy to follow.** The convergence guarantee for the gradient information matrix (10) is practically not possible because we can not a-priori draw samples from the posterior but only use samples pushed from the prior to posterior, whose convergence (to the posterior) analysis is beyond the current work. The complexity/scalability in # data points is actually shown in the middle of Figure 4 for pSVGD. Thank you for the suggestion to study the scalability w.r.t. the data dimension, which may lead to development of a new algorithm of projection in data dimension based on the low rankness of the data correlation matrix, which will be further explored. We will address other minor issues.