We thank the reviewers for their time and valuable feedback. We are happy to see that the four reviews are positive and describe our paper as well-written, clear, technically correct and interesting. We provide detailed responses below.

Reviewer #2: We thank the reviewer for giving a positive evaluation and valuable comments.

"Could you elaborate how the Lipschitz assumption can be relaxed?" The Lipschitz assumption is only used for bounding the integrand in (S28) in the supplementary document, as demonstrated in the derivations below line 180 (supp. doc). By assuming the uniform continuity instead, we can adapt these derivations to obtain another bound, but this requires the introduction of additional assumptions on the modulus of continuity. In that sense, we acknowledge that our remark might be confusing and decided to remove it from the main text to make the paper more self-contained.

"One major advantage of the sliced divergences is the dimension-independent sample complexity. But is it meaningful to compare the sample complexity of a sliced divergence and its based divergence?" This is a fair question and the reviewer is right in saying that the sliced divergence and its base divergence might have *"different scales"* based on Corollary 1. The main focus of our paper is to show that slicing results in a dimension-free convergence rate while carrying out a lot of useful topological properties of the base divergence (e.g. metric axioms, weak convergence). If the focus is on sustaining such topological properties, then we would argue that the improvement in the convergence rate is indeed meaningful. On the other hand, slicing also results in less discriminant divergences, as we mentioned for the IPM case (line 161), and in such a case, the improvement in the rate might be less significant. More analysis is required to understand the potential reduction in the discriminative power, and we leave it out of scope of this study. We will add a discussion about this point in the manuscript, and we will also make the suggested changes to the statement in line 53.

"Proposition 2 can be sharpened using the concentration of measure on the sphere" We are grateful to the reviewer for explaining this technique, which indeed leads to an improved bound. We will update Proposition 2 accordingly.

"Are there clean examples where $S\Delta(\mu_n, \nu_n) \to 0$, but $\Delta(\mu_n, \nu_n)$ -/-> 0 for TV distance or for other divergences with unbounded domain?" We are not aware of such example. We plan on further investigating this non-trivial question as future work, since this might help understanding whether Theorem 3 can be extended to non-compact domains.

Reviewer #3: We thank the reviewer for the positive evaluation and constructive comments. We will include the suggested additional references in our paper and clarify their connection with our contributions.

"When comparing distributions, projections onto some directions are arguably more "important" in terms of distinguishing distributions. will this improve the results?" We agree with the reviewer that sampling the projection directions uniformly on the sphere is not an optimal choice. However, we underline that this is the most common method used in practice, and this is why our paper relies on this technique. The study of sliced probability divergences based on non-uniform distributions is actually a very recent research topic, which raises interesting new challenges [1]. In that sense, investigating the consequences of the sampling scheme on our results, especially on Theorem 2 and the projection complexity as suggested by the reviewer, is a great idea: according to our derivations below line 80 in the supplementary document, Theorem 2 holds for *any* density $\boldsymbol{\sigma}$ defined on the unit sphere ; the bound in Theorem 6 illustrates the effects of the sampling distribution on the Monte Carlo approximation error, and might help tuning this distribution.

Reviewer #4: We thank the reviewer for their positive feedback and evaluation. *"While I imagine the result would be interesting and important for a subcommunity of NeurIPS, IMHO it looks more like it belongs in a math journal"* We believe that our paper is a good fit for NeurIPS, since our theoretical contributions revolve around metrics that form the basis of several computational statistics/machine learning methods. Understanding the performance of practical algorithms by providing theoretical guarantees is an important task in machine learning. In particular, the level of technicality of our work is very similar to the recent literature on similar topics [2, 3, 4], which were all published at NeurIPS 2019 and AISTATS 2019.

Reviewer #5: We thank the reviewer for their highly positive evaluation. As suggested, in order to increase clarity, we will revise the statement of Theorem 1 and move the explanation for the notation θ^* next to the definition of the Sliced-Wasserstein distance.

References

[1] Khai Nguyen, Nhat Ho, Tung Pham, and Hung Bui. Distributional Sliced-Wasserstein and applications to generative modeling, 2020.

[2] Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of Sinkhorn divergences. In *Proceedings of Machine Learning Research*, volume 89. 2019.

[3] Kimia Nadjahi, Alain Durmus, Umut Simsekli, and Roland Badeau. Asymptotic guarantees for learning generative models with the Sliced-Wasserstein distance. In *Advances in Neural Information Processing Systems 32*. 2019.

[4] Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems 32*. 2019.