We would like to thank the reviewers for their efforts on evaluating our paper. We appreciate that they pointed out the importance of the problem, and carefully checked our analysis. All the minor comments have been corrected as suggested, and we commit to releasing the source code around september. We address some common concerns below, and focus next on specific answers to each reviewer.

A. Empirical evaluations: Our current working draft has significantly improved numerical results. We cannot put them into this one-page response, but describe them succinctly. We have conducted experiments on Wasserstein barycenter problem for large document datasets with word embedding and report the results. Second, we add comparisons with other baselines (see **Point B** for details). Finally, we will follow the recommendations of R3 in terms of wording.

B. Baseline approaches: We thank R3 and R5 for pointing out that ADMM, ProxIBP, APDAGD and accelerated IBP should be included in the experiment. ProxIBP is a combination of IBP and proximal point method. The same strategy can be applied to FastIBP so it seems better to compare ProxIBP and ProxFastIBP. We will include those comparisons in the revision. For ADMM, APDAGD and accelerated IBP, we tried to implement them but found the performance is not ideal. This might be due to implementation issues. For example, APDAGD and accelerated IBP both require the line search procedure which is sensitive to parameter tuning. We will report our findings in more details in the revision.

C. Extension to more general settings: We can certainly extend our results without assuming that all measures are supported on the same number of points. We assumed it to avoid heavier notations. It is true that the ideas of this paper can not be generalized to the free-support Wasserstein barycenter problem. Indeed, the computation of free-support barycenters requires solving a multimarginal OT problem where the complexity bounds of algorithms become much worse; please refer to [2] for details.

In the following, we provide answers to specific questions raised by each referee.

Reviewer 1. We thank R1 for your evaluation. Please see **Point A** for empirical evaluations.

Reviewer 3. We thank R3 for your comments. Please see **Point A** for empirical evaluations and **Point B** for baseline approaches.

♦ *Intuition behind steps 1 and 3 in Algorithm 2.:* These steps are inspired by the momentum step in Nesterov's accelerated gradient method for convex smooth optimization. While the convergence property of IBP is only built on the convexity of the dual objective function [1], steps 1-3 further exploits the smoothness of φ to achieve the fast rate: $\varphi(\check{\lambda}^t, \check{\tau}^t) - \varphi(\lambda^\star, \tau^\star) = O(t^{-2})$. This is a crucial step, which leads to our improved complexity bound in Theorem 4.3.

Reviewer 4. We thank R4 for your review. Please see **Point A** for empirical evaluations.

♦ *The proof of Theorems 3.5 and 4.3 are in the supplement.:* We will add two detailed remarks behind Theorems 3.5 and 4.3 to further elaborate the proof ideas in the main context of the revised version.

Reviewer 5. We thank R5 for your evaluation. Please see **Point A** for empirical evaluations, **Point B** for baseline approaches and **Point C** for the extension of our approach.

♦ *The writing needs to improve, as it feels like much was left out and unclear in the shortening of the paper.:* We agree. We will reorganize the paper by spelling out the connection between Sections 2.3 and 4 explicitly. We will also revise the confusing parts in Section 3 as you suggested to improve its readability.

♦ *The algorithm is not tested in regimes of varying n to see how the algorithm scales compared to other existing algorithms.:* We have tested in regimes of varying n and presented the results in the supplement material; see Table 2.

♦ *why can* FASTIBP *achieve such bound?:* We will clarify in the updated version that FASTIBP is not a trivial combination: (i) steps 1-3 and 8 are standard in optimization literature yet first introduced to accelerated OT algorithms; (ii) steps 4-7 are specialized to the fixed-support barycenter problem. Furthermore, accelerated IBP and APDAGD are based on primal-dual framework, which allows for directly optimizing E_t and thus achieves better dependence on $1/\varepsilon$ than FASTIBP. In contrast, FASTIBP optimizes E_t through the dual objective gap, which leads to better dependence on n. Investigating the relationship between these two frameworks is interesting and we will try to provide discussions in the revision.

References

[1] S. Guminov, P. Dvurechensky, N. Tupitsa, and A. Gasnikov. Accelerated alternating minimization, accelerated Sinkhorn's algorithm and accelerated iterative Bregman projections. *ArXiv Preprint: 1906.03622*, 2019.

[2] T. Lin, N. Ho, M. Cuturi, and M. I. Jordan. On the complexity of approximating multimarginal optimal transport. *ArXiv Preprint: 1910.00152*, 2019.