We thank the reviewers for their time and their valuable feedback and thoughtful suggestions. We remain confident that our work is of strong interest to the NeurIPS community and can be further strengthened by incorporating the suggested changes in a final version of the paper. Below are our answers to the reviewers' comments, grouped by topic.

Novelty of dual formulation. **R2** considered our method requiring a "discretized proxy." This is not correct. The proxy support measure in the formulation is a continuous measure, and in the implementation we choose a uniform measure on the bounding box containing the support of all source distributions (estimated by samples). Such choice of support measure allows us to design a novel barycenter algorithm in a *free-support* and *continuous* setting: both the source distributions and the resulting barycenter are continuous (except for optional recovery method (a), but we suggest continuous alternatives). To the best of our knowledge, all prior works either consider *fixed-support* (e.g., [3]) or *discrete* source measures (e.g., [2]). We thank **R2** for pointing out [3] which we will add to the references. The method in [3] utilizes entropic regularization and uses a fixed discrete proxy, making it sufficient to parameterize only one of the dual potentials as the other can be retrieved in closed form based on first-order optimality [3]. For a continuous setting, this becomes infeasible as the closed form will become an integral in high dimensions. Moreover, such closed form is not possible for quadratic regularization which we found to be more stable and more accurate in higher dimensions.

R2 and **R3** asked about the novelty of our approach over existing continuous methods on regularized OT distances [4, 5]. First of all, a different, more challenging optimization problem is studied in our work. The variables in the barycenter problem we consider include not only the individual transport plan from each source to the barycenter, but importantly also the barycenter itself. There is no clear way to extend [4, 5] to the barycenter setting without introducing non-convex min-max optimization. By introducing a novel regularizing measure $\mu_i \otimes \eta$ that does not rely on the unknown barycenter but only on a proxy measure η, we are able to encode the information of the barycenter in the dual potentials themselves without an explicit parametrization, thanks to Theorem 3.1. We will further emphasize the novelty of our method compared to [4, 5] in the final version of the related-work section.

Comparison of recovery methods. We agree with **R3** that the numerical comparison of 5 barycenter recovery methods is another highlight of the paper; even for computing regularized transport distance such comparison is lacking in the existing literature. In particular, we proposed the MCMC recovery strategy (b) which is suitable theoretically since we use continuous parametrization of the dual potentials (so gradient-based MCMC methods are viable) but is slow in practice. To address **R2**'s concern of lack of conclusion on which method to choose, we have included a discussion explaining each method's advantages and disadvantages in the paragraph "qualitative results in 2 and 3 dimensions" on page 6. We will add recommendation for higher-dimensional situations to accompany the results in Table B.1 and Table B.2. To address **R1**'s concern about recovery methods being costly, method (d) in fact comes at almost no cost (gradient computation), and method (e) requires a secondary SGD optimization that is no more costly than Algorithm 1.

Relevance to NeurIPS. **R3** suggested applying similar strategies to a wider range of variational problems of interest to NeurIPS. We thank **R3** for the detailed references, which we will incorporate into the related-work section. We would like to point out that there are three accepted papers at NeurIPS last year inspired by Wasserstein barycenters. The application to large-scale Bayesian posterior computation is also of considerable interest to the NeurIPS community.

Theoretical analysis. **R3** suggested analyzing the duality gap for the algorithm in practice since parametrization will restrict the functional space. On a related note, **R1** asked about theoretical analysis of recovery methods. These are challenging questions that depend on the specific structure of parameterization and the particular recovery method. Even for continuous regularized OT distance [5] this is not well-understood. We agree these are areas for future work.

Experiments and improvements over SOTA. **R2** is concerned about the level of improvements over SOTA. Our improvements are quite significant in the experiments: in Table 1, the covariance difference computed by our algorithm is consistently 1.5-2$\times$ smaller in higher dimensions as those of SOTA. Additionally, this example only deals with Gaussians, whose low sample complexity favors discretized methods more than other settings. **R2** mentions the efficiency of [6], but [6] is not applicable in high dimensions as a discretized grid is needed. While improvements in Table 2 appear more modest, covariance difference does not fully characterize discrepancy of measures for non-Gaussians. The comparison using 2-Wasserstein distance in Figure 3 reveals significant improvement, which we expect to further increase if we use more samples from our barycenter for the 2-Wasserstein distance. **R4** suggested testing with different MLP architectures. We tested with deeper/wider networks but found no noticeable improvement over the proposed architecture. We will add more details on the runtimes for different methods as **R4** suggested. In high dimensional experiments, our method takes around 15 minutes, [2] takes 20 minutes, and [1] takes an hour or longer. The relative ease of training is likely due to the convexity of our formulation (Equation 11).

[1] Claici et al. Stochastic Wasserstein barycenters. *Proc. ICML*, 2018.

[2] Cuturi & Doucet. Fast computation of Wasserstein barycenters. *J. Machine Learning Research*, 2014.

[3] Cuturi & Peyré. A smoothed dual approach for variational Wasserstein problems. *SIAM J. Imaging Sciences*, 2016.

[4] Genevay et al. Stochastic optimization for large-scale optimal transport. *Proc. NeurIPS*, 3440–3448, 2016.

[5] Seguy et al. Large-scale optimal transport and mapping estimation. *Proc. ICLR*, 2018.