We would like to thank the meta reviewer and three reviewers for the care with which you handle the submission and for your professional and constructive comments. We have made every effort to address the concerns.

Response to Reviewer 1's comments

1 In the revision, we will review and discuss more OT approaches that are not based on solving a linear program.

2 Our algorithm utilizes a second-order dimension reduction method to estimate the projection direction. Hence, the leading projection direction corresponds to the direction of the maximum marginal discrepancy between the **variances** of the distributions. Our paper and some dimension reduction literature made bad presentations by ignoring third and higher-order moments. We apologize for this ambiguity and will make it clear in the revision.

3 (1) In the revision, we will add dimensions to the quantities in algorithms. Yes, we require $2n > d$ for the reason you mention. A viable stopping criterion of Algorithm 2 is to check the angle difference of projection directions between two consecutive iterations. The algorithm is terminated when the angle is close to zero. We appreciate these constructive comments. (2) Yes, the computational cost analysis follows the assumption that $d \ll n^{2/3}$, which allows us to dominate $O(d^3)$ by $O(n^2)$. The lookup table in Algorithm 2 is simply sorting. In the revision, we will explain it clearly and review more literature for the 1d optimal transport and lookup table method. (3) We will revise Assumption 2 with population-level language. We require Assumption 2 (c) to hold for a fixed integer $r > 1$, not every positive integer. We have fixed the typo in Line 196. We acknowledge the enlightening paper you recommend.

4 (1) The non-monotonic convergence is caused by the non-equal sample means of two point clouds which can cause some troubles to violates the assumptions of SAVE. A remedy is to use a first-order dimension reduction method like SIR to adjust means first. We find it empirically solves the problem. In our experiments, we observe that RANDOM outperforms SLICES in simulations but vice versa for real data. This may suggest that RANDOM is more greedy but less robust. In light of your suggestion, we will report the variance over replications. (2) We do not observe the over-fitting of PPMM in our experiments. The projection direction found by Algorithm 1 tends to converge when $\mathbf{X}^{[k]}$ converges. In contrast, we do observe that RANDOM and SLICED deteriorates in some scenarios.

5 IMPROVEMENTS: (1) In Algorithm 1, the estimation accuracy of $\widehat{\Sigma}$ depends on the tail probability of X and Y. Also, the OTM estimator in Algorithm 2 will be affected by asymmetry and outliers. So we expect the algorithms to perform best when X and Y follow symmetric and sub-exponential distributions (e.g., Gaussian). (2) To converge fast, we require the eigenvalues of Σ to decay fast enough (approximately low rank). Hence, the first a few projection directions can explain the majority part of the variance of the discrepancy between two distributions.

Response to Reviewer 3's comments

6 Thank you for this insightful comment. Our method can be extended to the cases of non-equal sample-sizes and non-equal weights with small tweaks. For two point clouds with non-equal sizes, we can use the approximate-lookup table in Algorithm 2. Also, we can calculate a weighted covariance matrix in Algorithm 1 to allow non-equal weights. Here we use a simulated example to demonstrate these cases. We follow a similar setting as in Section 4.1 except that we draw $5,000$ and $1,000$ points from p_X and p_Y, respectively. We set $d = 10$ and assign weights to observations randomly. The results are presented in Fig. 1, where the colored lines are the sample means of estimated Wasserstein distances over 100 replications and the black dashed line is then calculated by the "short simplex" method, which serves an oracle. In addition, the average wall time (until converge) is: PPMM(0.3s), RANDOM (1.4s), SLICED10 (14s) and "short simplex" (74s). In the revision, we will discuss this important extension with additional numerical justifications. We believe that such an extension will make the proposed algorithm applicable to a much larger family of OT problems. We also plan to discuss the extension of the algorithm to Kantoronovich's formulation.

7 (1) The conditions in Assumption 1 are widely used in dimension reduction literature and are not as restrictive as they seem. It will not prohibit a nonlinear model. Intuitively, (a) and (b) assumes that $u^T Z$ behaves like Normal. (2) We agree with your comment on complexity statements and will re-write this part accordingly. Yes, ϕ is a permutation, and the lookup table step is just sorting. We will make corrections according to your minor comments. In Line 202, O_p stands for order in probability which is similar to O but for random variables. (3) In the revision, we will make a comprehensive discussion of assumptions and how they affect the convergence analysis.

Response to Reviewer 4's comments

8 Our algorithm can be extended to address your concern about non-equal sample-sizes and non-equal weights. Please see our response 6 to Review #3 for more details and a simulation result. We appreciate this important comment.

9 (1) In the revision, we will report the wall time. (2) Also, we will compare our method with fast transport solver and standard perfect matching. (3) The paper from Paty and Cuturi is very interesting and helpful. We will cite and discuss it in the revision. (4) The memory cost is $O(nd + d^2)$. We will discuss it in the revision. We are grateful for these constructive comments.

10 The assumption in line 177 is for mathematical simplicity. It can be removed if we use a first-order dimension reduction method like SIR to adjust means before we apply SAVE. Our theory can be readily applied to this case. We thank the reviewer for pointing out this insightful issue.

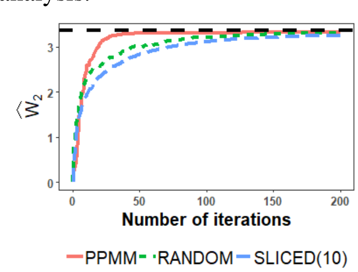

Figure 1: Simulation with the point clouds with different sample-size and non-equal weights