We kindly thank the reviewers for their positive comments and insightful suggestions. Detailed responses below.

Reviewer 1:

- *"I was not entirely sure why we would expect R to appear in the runtime? Any intuition, e.g. whether it is inevitable or is an effect of this analysis, would be useful."* This is a good question for which there is an intuitive answer: the dependence on the radius R is inevitable for any algorithm which approximates OT to ε additive error, essentially due to a scale-invariance argument. Specifically, since the transportation cost $\sum_{ij} P_{ij}\|x_i - x_j\|_2^2$ is part of the definition of the Sinkorn distance (see eq (1) in the paper), if all pairwise distances $\|x_i - x_j\|_2^2$ increase by factor of say 10, then so does the transportation cost. In other words, approximating Sinkhorn divergence to ε additive error becomes harder as the radius R increases. We will add a comment to the camera-ready version to clarify this.

Reviewer 2:

- We will be sure to revise the camera-ready version to be consistent with those terminologies.
- We will add a comment about the applicability of Nyström to other, more general kernels. However, the approximation guarantees in our analysis (in Section 3) are specifically tailored to the Gaussian kernel.
- *"Adaptive Nyström is fine but .... memory/runtime consumption.* Yes, definitely agreed. This is precisely the purpose of the guarantees we proved in Section 3, which give provable guarantees on the number of landmarks required to obtain a given approximation error of the kernel matrix. These guarantees are then used to prove Theorem 1 (our main result) which gives provable bounds on the memory/runtime consumption of our proposed algorithm Nys-Sink.
- *"Could you please also comment on the out of sample extension... would it be 'safe' to calculate the proximities exact?"* Yes, out-of-sample extension is totally doable with Nyström.
- *"It would be nice to have a larger evaluation on the practical impact (not only runtime) of the approximation on its usage in some algorithm (so far only Fig 2) –> this is somewhat covered in the supplement but its a bit sad not to have a clear summary in the main paper."* We will move some of the relevant material from the supplement, and make sure to clarify in the main text.

Reviewer 3:

- We will be sure to revise the camera-ready version to define those notations, and to make clearer the background on the Sinkhorn distance.
- The reason we only consider squared Euclidean cost is that it is only in this case that the corresponding kernel matrix K is a Gaussian kernel matrix, which our algorithms and analysis heavily rely upon. Nevertheless, we emphasize that in many applications, especially in the field of computer graphics, image processing and simulation of physical systems, the 2-Wasserstein distance is a very common choice. It is also among the most well-studied choices in theory. Our focus on this distance is therefore of both theoretical and practical significance.