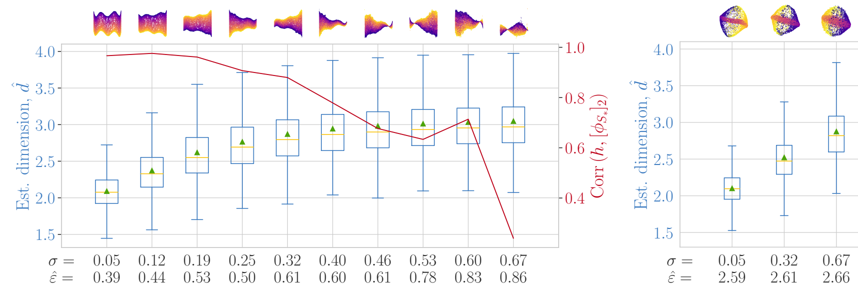
Thanks for the VERY careful, responsible and competent reviews our paper has received! We will implement all improvements recommended in the 3 reviews. Here we comment only on the more significant questions raised.

Reviewer 1 *" relate to: "Non-Redundant Spectral Dimensionality Reduction", Michaeli et al."* Will do. Thanks for pointing us to this reference. *" The choice of kernel bandwidth (ε) not addressed."* For the real data, ε was optimized as in [JMM17]. For the synthetic data, ε was chosen heuristically; since, experiments were rerun using [JMM17] (see also below). *"if ε is chosen as a diag matrix…, the aspect ratio problem could be fixed (see for example "Kernel Scaling for Manifold Learning and Classification"). To summarize, I think the paper should be accepted and hope that these minor changes could be easily addressed to improve this manuscript."* We will discuss this reference in final paper.

Reviewer 2 *"…experiment on synthetic data with added noise"* Experiment with the 6.28×2 strip data (below, left): Gaussian noise with standard deviation σ and ambient dimension $D = 3$ was added; for each σ, the ε selection algorithm [JMM17] was run, as well as the IndEigenSearch algorithm for selecting the coordinate for embedding in the top row, and intrinsic dimension estimation [LB04]. $\hat{d}$ measures the degradation of the manifold structure due to noise, and Corr the recovery of h (shorter dimension in stripe). We see that IndEigenSearch degrades little even when $\hat{d} \approx 2.75$. Similar experiment on tall torus is below, right. In the submission, $\sigma = 0.05$ and the heuristic ε was 0.25 for stripe and 1.5 for tall torus. *"…more interpretation of the utility of the embedding."* For MD data, the embeddings represent "slow motions" of the molecule (e.g., rotations of one group w.r.t. another); for galaxy spectra, it is interesting to compare Fig. 3.f. with the "HR diagram principal sequence", where stars align in spectral/brightness space in 1D, according to their ages. For galaxies, age of star population is also a feature, but the manifold is 2D. We now also have experiments with similar good results for UMAP embeddings initialized by coordinate sets chosen by IndEigenSearch.



Reviewer 4 *" the paper does not focus on how to optimize this objective function"* In a longer paper, optimization will receive more space. See also below, and Supplements C, D, E1. Note that for the current data sets, the run times for [JMM17]/DiffMap/IndEigenSearch are approximately in the ratio 30/3/1 (synthetic) and 100/10/1 (real).

" the IndEigenSearch problem chooses a composition of the original map with a very specific Euclidean projection: a projection along coordinate axes. … Why is [searchign over sets better] than to search over all projections, ([by] e.g. manifold optimization on the Grassmannian)?" **This is a super-interesting question for future work, and we thank the Reviewer for raising it.** Presently, we can say that: the loss $\mathcal{L}(S)$ extends in a straightforward way to the Grassmanian manifold; $\mathcal{L}(P)$, with P a projection matrix, is a difference of convex functions, while the original $\mathcal{L}(S)$ is a difference of *submodular functions* – see Supplement. **Computational aspects:** for small s or m, there are only ~ 200 $\mathcal{L}$ calculations; the search for S is insignificant compared to computing the embedding (in particular, the neighborhood graph and ε search). When m, s grow, the brute force IndEigenSearch cost will grow exponentially. The user has the choice between more advanced discrete optimization over S, based on submodularity, vs continuous optimization over P, but of essentially the same function. A minor but nice advantage of searching over sets is that it only requires the manifold learning toolbox; a practitioner needs not get tools (e.g. `manopt`) for optimization over the Grassmanian manifold.
Mathematically, however, the question is deep and significant: can there be an advantage in using a linear combination of eigenfunctions, instead of a subset? More specifically, for manifolds with small injectivity radius and large aspect ratios, could it be that the required embedding dimension s is smaller if we optimize over the Grassmanian and not over discrete subsets of coordinates? We did not find any answers to this in the literature (so far).

"…why is K-L between these two volume forms a good way to encourage local injectivity." Local injectivity is by definition tied to a volume form j (sorry for yet another unusual notation); the only question is how do we "compare it with 0". We compare it with its maximum $\tilde{j}_S$; then we integrate over the "inability to reach the max", which is exactly what a K-L divergence does. Stretching it some, pj_S is the "data" and $p\tilde{j}_S$ is the "model", and we are looking for a view S of the data that agrees with the model. Here p is the density of the data sampled from a distribution on $\mathcal{M}$, see also Assumption 2 in the manuscript.

References

[JMM17] "Improved graph Laplacian via geometric self-consistency" by Joncas et al., NeurIPS 2017
[LB04] "Maximum Likelihood Estimation of Intrinsic Dimension" by Levina and Bickel, NeurIPS 2004.