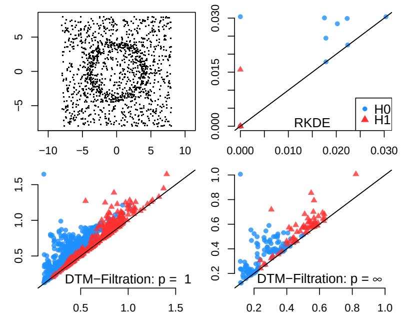
We thank the reviewers for their time and effort in providing insightful and invaluable feedback. Below, we first address the major common concerns of the reviewers, and then their individual comments. All minor comments will be addressed in the camera-ready version. Additionally, all code related to experiments will be made publicly available.

(I) Computational time. In our experiments, the computational bottleneck is the persistent homology pipeline, and not the KIRWLS for $f_{\rho,\sigma}$ (RKDE). A simple runtime analysis is presented in Table 1. For $n = 1000$, $\mathbb{X}_n$ is sampled from a torus in $\mathbb{R}^3$, and the *total* time taken to compute the persistence diagrams (PDs) is reported for several grid resolutions.

(II) Comparison to DTM-Filtrations. We highlight some differences between our approach and those in [A, B]. First, as remarked in [A, §5], most properties of the DTM-Filtration follow from the stability of DTM w.r.t. Wasserstein metric. Similar to DTM, it can be shown that the robust KDEs (and KDEs) exhibit stability w.r.t. MMD metric (Maximum Mean Discrepancy). However, it should be noted that stability is inherently different from robustness, which we have expounded in our analysis using the *persistence influence*. In this context, the figure indicates the advantage of our proposed approach in the presence of adverse noise. Second, we note that we use superlevel filtrations in the experiments, in contrast to weighted Rips filtrations (which is computationally appealing, especially in higher dimensions). Notwithstanding, extending our proposed approach (i) to power-distances for constructing weighted Rips filtrations using the ideas in [29, §3] and [C, Chap. 5], and (ii) using coresets, as in [B], will be interesting future directions. Finally, we would like to emphasize that the objective of our work was to illustrate that outlier-robust persistence diagrams can be constructed without compromising on statistical efficiency, with the hope that the theoretical tools presented here serve as a stepping-stone for developing *efficient and robust* PDs. We will incorporate and expand on this discussion in the revised version.



(III) Additional experiment. We perform a variant of the six-class benchmark experiment in [1, §6.1] to address some concerns shared by $\mathcal{R}\#1$–3. 25 point clouds are sampled from each of six 3D "objects" with additive Gaussian noise ($\sigma = 0.1$), *and* ambient Matérn cluster noise. $\mathsf{Dgm}\,(f_{\rho,\sigma})$ is the PD constructed using $f_{\rho,\sigma}$; and $\mathsf{Dgm}\,(d_{\mathbb{X}_n})$, from the distance function $d_{\mathbb{X}_n}$, is transformed to the persistence image $\mathsf{Img}\,(d_{\mathbb{X}_n})$. Note that the former is a robust PD while the latter is a stable vectorization of a non-robust PD. Spectral clustering is performed on the resulting distance-matrices: W_p metric for $\mathsf{Dgm}\,(f_{\rho,\sigma})$, and L_p metric for $\mathsf{Img}\,(d_{\mathbb{X}_n})$. The quality of the clustering is assessed using the rand-index. The results are reported in Table 2. We will include a detailed version of this experiment in the revised version.

Table 1.	0.04	0.06	0.08	0.10	Grid-size
RKDE Dgm	76.7	17.1	6.7	3.5	Runtime
KDE Dgm	75.5	15.3	4.7	1.8	(in Seconds)

Table 2.	W_1/L_1	W_2/L_2	W_∞/L_∞
H_0, $\mathsf{Dgm}\,(f_{\rho,\sigma})$	0.9093	0.9280	0.9032
H_0, $\mathsf{Img}\,(d_{\mathbb{X}_n})$	0.8612	0.8684	0.8723

Response to $\mathcal{R}\#1$: Please see **(II)** and **(III)** regarding the comparison with [A] and experiments. **1.** We agree that the claim on `L.29-30` is questionable, and it will be removed; our intention was to highlight that there are two similar (but different) approaches. **2.** We agree that the space has curvature unbounded from above, *and* bounded from below [D, Thm 2.5]. **3.** We use ϕ(`\phi`) for filter functions and φ(`\varphi`) on `L.69`. **4.** $\phi_{\mathbb{P}^\epsilon_{\boldsymbol{x}}}$ is the filter function induced by $\mathbb{P}^\epsilon_{\boldsymbol{x}}$ on `L.167`. **5.** As the final version allows an additional page, we will have more space to alleviate the denseness. We will carefully include all the necessary terminology as well as a concise introduction to kernel theory for enhancing clarity.

Response to $\mathcal{R}\#2$: Please see **(II)** and **(III)** regarding the comparison with [A, B] and experiments in 3D. **1.** The numerical implementation is done using cubical homology; this will be clarified in the revised version. Despite being infeasible in very high dimensions, this method is still widely used in applications (e.g., [40]). **2.** The runtime analysis is discussed in **(I)**. The quality of the output can deteriorate if the grid is too fine [19, Lem 11]. **3.** The concern related to `L.29-30` is discussed in **1.** for $\mathcal{R}\#1$. **4.** $\mathcal{D}_\sigma$ is the space of mean embeddings, and is defined in `L.65`. **5.** ℓ is not pre-defined, but the KIRWLS algorithm is run until the relative change of empirical risk is less than 10^{-6}. In practice, we have observed ℓ to be well below 100. **6.** Thm 4.1 only uses $(\mathcal{A}1)$–$(\mathcal{A}3)$, and a part of $(\mathcal{A}4)$ is used in Rmk 4.1. **7.** Rmk 4.1 (i) is meant to say "*a similar bound holds when ...*", and $\alpha < 2$ in (ii). **8.** *Proof*: Lem B.1 will be added and renumbered, as suggested. We also agree that the existence assumption must be included in the statement to Thm 4.1. Thank you for pointing this out, and for the insightful comment on the implication of Thm 4.1 for linear vectorizations of robust PDs. **9.** Replacing W_∞ by W_p and using the W_p stability theorem in [E, §5] ensures that Thms 4.2 & 4.3 still hold, while Thms 4.1 & 4.4 will require a careful analysis. We will include these clarifications in the revised version.

Response to $\mathcal{R}\#3$: Please see **(I)** regarding the empirical time; we will include this information in the revised version. We also address the concern related to uniform noise in **(III)**, where we consider noise from a Matérn cluster process. The implication of using W_p instead of W_∞ for examining persistence influence is discussed in **9.** for $\mathcal{R}\#2$.

Response to $\mathcal{R}\#4$: Please see **(I)** regarding the concern about computational time. In the final version, we will report the runtime for all experiments. Regarding the concern of reproducibility, we will make the codes publicly available.

References. [A] Anai et al., 2018, SoCG; [B] Brécheteau et al., 2018; [C] Jisu Kim, 2018, PhD Thesis;
[D] Turner et al., 2014, DCG; [E] Cohen-Steiner et al., 2010, FoCM;