We thank the reviewers for their careful reading and valuable comments, which we address one by one below.

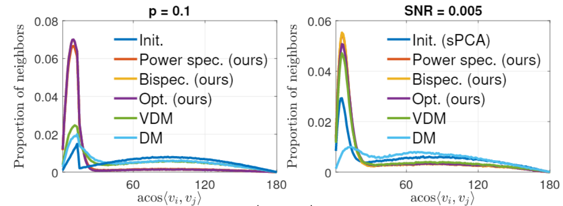

Figure 1: Histograms of $\arccos \langle v_i, v_j \rangle$ between identified nearest neighbors. *Left*: simulated data for $\mathcal{M} = \mathrm{SO}(3)$ under random rewiring model; *Right*: cryo-EM images. The clean histogram should peak at small angles.

Reviewer 1: Q1: *Show comparisons to other manifold learning methods, additional metrics, and classified objects.* **R:** We add the comparisons with Diffusion maps (DM) (see Fig. 1) and Laplacian eigenmaps (LE). The performance of LE is similar to DM, since the data are uniformly distributed on the manifold. In the first two experiments in Sec. 5 of our paper, we focus on the accuracy of the nearest neighbor identification given extremely noisy initial graph structure. Fig. 1 shows the geodesic distances of the estimated nearest neighbors on S^2. We use the **Jaccard index** as an additional metric. The Jaccard index evaluates the similarity of the estimated and the true nearest neighbors of each node. For the synthetic data on S^2 at $p = 0.1$, the mean Jaccard indices are *0.196 (Power Spec.), 0.209 (Bispec.),* ***0.215*** *(Opt.), 0.059 (VDM), 0.042 (DM)* (higher the better). For the cryo-EM images (SNR$= 0.005$), they are *0.033(Power Spec.),* ***0.035*** *(Bispec.), 0.031 (Opt.), 0.028 (VDM), 0.024 (DM).* We will add these results in revision. The histogram contains the information on how close the estimated nearest neighbors are, whereas the Jaccard index only measures the set similarity. For the application in spectral clustering, we use **rand index** to measure the performance in the paper. The F-score does not apply here since the examples are not binary (or multi-class) classification problems with labels. We will clarify the choice of metrics for the performance evaluation in revision. We will add more illustrations to show samples of estimated nearest neighbor images and improvement in image denoising (see response to R2 Q3 and Fig. 2).

Improvements: *More comprehensive evaluation of performance especially on real data.* **R:** We will add additional metric, comparisons, and illustrations mentioned in the response to R1 Q1 in the revised manuscript.

Reviewer 2: Q1: *Small typos and grammatical errors, m_k, direct sum.* **R:** Thanks for pointing these out, we will correct the typos/errors and clarify the definition of parameters in our revised manuscript.

Q2: *Tunable parameters?* **R:** The choice of parameters was explained in the captions of Figs. 2 and 3 of the paper; we will discuss them in greater detail in the revised version of the main paper. The maximum frequency k_{max} is chosen to be as large as possible within our computational budget—this is because it is empirically observed that the performance improves as k_{max} gets larger, but saturates once k_{max} becomes sufficiently large. The parameter for the number of eigenvectors m_k is chosen relatively small (≤ 50) for computational efficiency and to exclude the noise-sensitive "high-frequency" eigenvectors. For nearest neighbor searching, the number of nearest neighbors is chosen to ensure a well connected sparse graph in the noise-free setting. For spectral clustering, the initial graph is given and fixed.

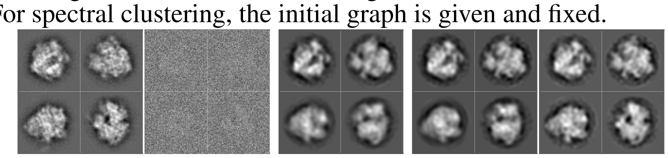
(a) Clean (b) Noisy (c) Init. (d) VDM (e) Bispec.

Figure 2: (a) Clean projections of 70S ribosome; (b) Noisy images with SNR $= 0.01$; (c) to (e) Denoised images based on the graph and alignments identified by the initial estimation, VDM and Bispectrum-like affinity (this paper). MSEs of the denoised images are (c) 6.24, (d) 5.72, and (e) **4.97** (lower is better).

Q3: *Application to the real cryo-EM data?* **R:** There is no direct way to compare the performance of nearest neighbor identification algorithms on real microscope images, since their viewing angles and underlying clean images are unknown. We used simulated data in our experiments so that the outputs can be compared and contrasted with the "ground truth." Nevertheless, as a proxy to real data experiments, we will add results demonstrating how the denoising step can benefit from the improved nearest neighbor identification (see Fig. 2); it is known that the quality of the denoised images directly contributes to the 3D reconstruction results (see e.g. explanation in reference [74] of the main paper).

Improvements: *Exposition and application.* **R:** We will move the algorithms into the supplement and add more explanations and intuition in the main paper. Our paper provides a framework for analyzing data that lie on or close to a manifold with a group action and is not limited to cryo-EM problem, e.g., spectral clustering with $\mathrm{SO}(2)$ and $\mathrm{SO}(3)$ transformations. We will add the cryo-EM denoising results in revision. Other tasks will be explored in the future.

Reviewer 3: Q1: *More background and intuition.* **R:** We will move Alg.2–4 to the supplementary material, and add more explanation on group theory and irreducible representations in the main paper. We will also provide motivating examples with $\mathrm{SO}(3)$ to explain the intuition of using Wigner D-matrices and Clebsch–Gordan coefficients.

Q2: *Gain of incorporating multiple representations over the "best" representation?* **R:** In practice, observations from real data—in any representation—always contain certain level of noise, even for the "best" representation. Incorporating multiple representations allows us to leverage the inherent consistency across different representations of the same information to better remove noise. Methodologically, incorporating multiple representations creates a "redundant" representation akin to redundant wavelets/frames/dictionaries in applied harmonic analysis, which are known to be more robust to noise due to the additional structural rigidity. We will further clarify this in the revised version.

Q3: *Clarify the applicability in cryo-EM and computer vision.* **R:** We will clarify in writing how the proposed work can be applied in cryo-EM and computer vision, and add more illustrations such as Fig. 2.

Improvements: *Remove Algs. 2–4. Write more about the background and intuition.* **R:** Please see response to R3 Q1.