Ab initio tomography with object heterogeneity and unknown viewing parameters
Supplemental Material

Arunabh Ghosh[1], Ritwick Chaudhry[2], and Ajit Rajwade[3]

[1]Dept. of Electrical Engineering, Indian Institute of Technology Bombay
[2]Adobe Research
[3]Dept. of Computer Science & Engineering, Indian Institute of Technology Bombay

Reproducible Research

The author's implementation of algorithms, example usage, and synthetic 2D protein datasets can be found in the Github repository Arunabh98/Tomography-Heterogeneity.

Patch-Based Denoising Algorithm

We use a patch-based PCA denoising method to reduce the noise in the projection clusters as mentioned in Sec. 2.2 and Sec. 4.1 of the paper. This algorithm is adapted from a similar algorithm for 2-D images, as described in [1]. In this algorithm, we extract small-sized patches from each cluster center. For each such patch, we find some L patches nearest to it in terms of the ℓ_2 distance. After performing PCA on this set of patches, we project each patch along the principal direction to produce eigen-coefficients. To denoise the patch, we manipulate these coefficients using Wiener-like updates of the form

$$\beta_{il} = \alpha_{il}\left(\frac{\sigma_l^2}{\sigma_l^2 + \sigma^2/\bar{K}}\right) \tag{1}$$

where β_{il} is an estimate of the l^{th} denoised coefficient for patch i (part of cluster center $\bar{p}_j$), α_{il} is the corresponding noisy coefficient, σ^2 is the noise variance in the original projections which is assumed to be known, $\bar{K}$ is the average number of projection vectors assigned to a cluster, and σ_l^2 is the mean square value of the l^{th} coefficient estimated as follows:

$$\sigma_l^2 = \max\left(0, \frac{1}{L}\sum_{i=1}^{L}\alpha_{il}^2 - \sigma^2/\bar{K}\right). \tag{2}$$

This patch-based PCA denoising algorithm is better than the PCA denoising algorithm used in [2]. This is because, in [2], entire projections are compared instead of just patches. The advantage of our patch-based approach is that it is easier to find a number of patches which are structurally similar to a given reference patch, but that is not true for entire projections. The second advantage is that we now have the freedom to find similar patches from within a projection vector, but from other projection vectors as well. Moreover, we performed the denoising in sliding window fashion with a pixel stride of 1, resulting in several potential denoised values per pixel. The final denoised value was selected using averaging. Using this denoising scheme, we remove the residual noise (averaging step) and thus obtain a representative set of significantly less noisy projections.

Evaluation

0.1 Single Linkage Clustering

To support our claim that the single linkage clustering algorithm generates clusters composed of projections from just one conformation we run the following experiment: A total of $Q = 3 \times 10^4$ projections subjected to noise

from $\mathcal{N}(0, \sigma^2)$ where $\sigma = 0.3a$ are taken from the conformations shown in Fig. 1 and are clustered using three standard clustering algorithms: K-means clustering, Single-linkage clustering and Centroid linkage clustering (a variant of agglomerative clustering where the distance between centroids is used as a metric). We measure the purity of a cluster as follows:

$$\tilde{c}_j = \frac{\sum_{p_i \in \chi_j} p_i(1 - I_j(p_i))}{N_j} \tag{3}$$

where $I_j(p_i) = 0$, when the i^{th} projection belongs to the majority class in the j^{th} cluster, and 1 otherwise.

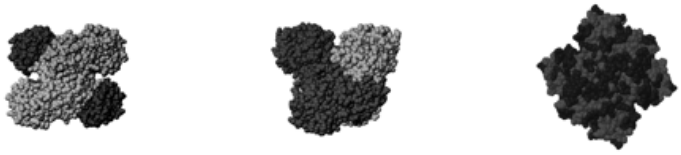


Figure 1: The three conformations of Holo-Glyceraldehyde-3-Phosphate Dehydrogenase

We measure the performance of the clustering algorithm by taking an average of the purity $\tilde{c}_j$ of all the clusters. Table 1 demonstrates the performance of each algorithm averaged over 10 runs of the experiment. As seen, the single linkage clustering generates clusters with high purity which can be used later for an accurate classification.

Clustering Algorithm	**Average Cluster Purity**
K-means clustering	81.09%
Centroid Linkage	82.44%
Single Linkage	99.78%

Table 1: Performance of different clustering algorithms.

0.2 Reconstructions

This section contains some additional results of the reconstruction of all the conformations of various proteins ranging from Lipase to Hemoglobin under varying levels of noise. The near-perfect reconstruction in each case demonstrates the broad applicability of our algorithm as well as its robustness. For each protein, we show the original conformations, the classification provided by the moments based approach (Sec. 3.1 of the paper), refined classification using the graph-Laplacian based approach (Sec. 3.2 and Sec. 3.3 of the paper) and final reconstructions along with the RMSE error. Note that in $\sigma \triangleq \beta a$, a is the average noiseless projection value.

0.2.1 Lipase

A total of $Q = 5 \times 10^4$ projections are simulated where each projection is subjected to additive i.i.d. noise from $\mathcal{N}(0, \sigma^2)$ where $\sigma = 0.35a$.

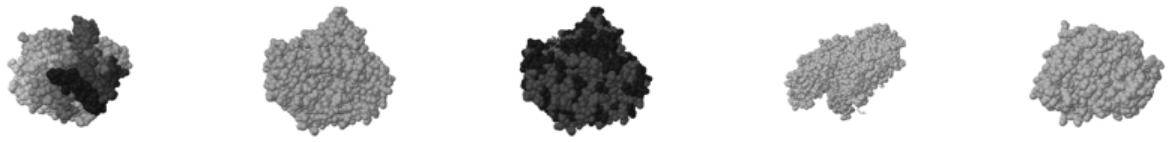
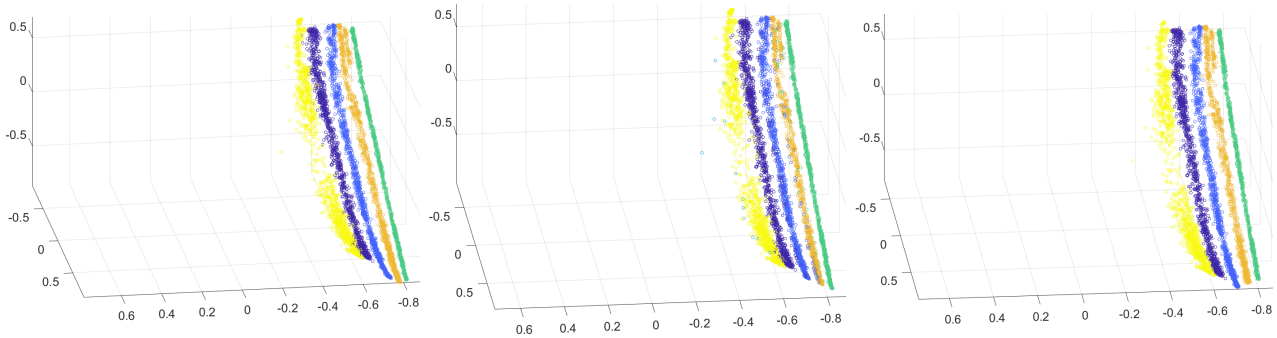
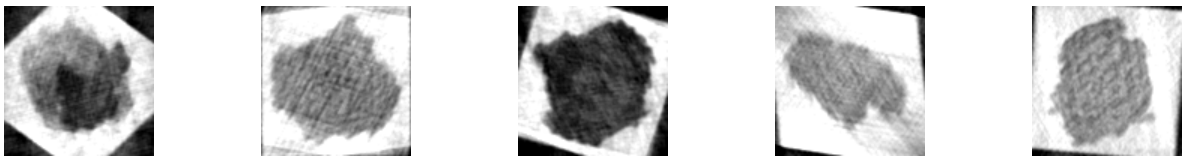


Figure 2: The original 5 conformations of Lipase are shown in this figure.



(a) Original clusters (b) Initial Classification (c) Final Classification

Figure 3: Left: The original ground truth classification of clusters. Center: The classification provided by the moments based approach. Right: Final refined classification.



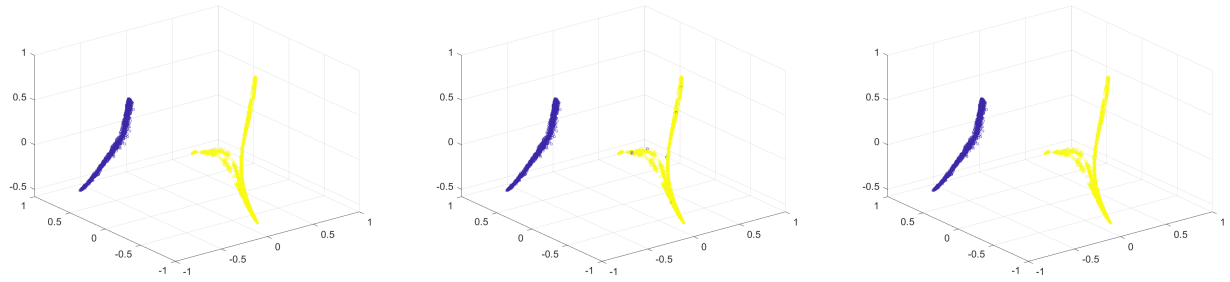
(a) 19.08% (b) 17.20% (c) 21.62% (d) 16.80% (e) 18.68%

Figure 4: The final structures estimated by our algorithm.

0.2.2 Taq DNA polymerase I

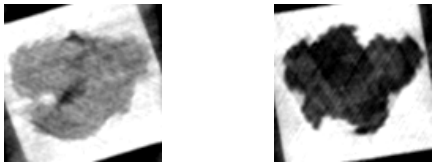
A total of $Q = 2 \times 10^4$ projections are simulated where each projection is subjected to additive i.i.d. noise from $\mathcal{N}(0, \sigma^2)$ where $\sigma = 0.3a$.

Figure 5: The original 2 conformations of Taq DNA polymerase I are shown in this figure.



(a) Original clusters (b) Initial Classification (c) Final Classification

Figure 6: Left: The original ground truth classification of clusters. Center: The classification provided by the moments based approach. Right: Final refined classification.



(a) 16.97% (b) 20.21%

Figure 7: The final structures estimated by our algorithm.

0.2.3 Guanylate Kinase

A total of $Q = 3 \times 10^4$ projections are simulated where each projection is subjected to additive i.i.d. noise from $\mathcal{N}(0, \sigma^2)$ where $\sigma = 0.3a$.

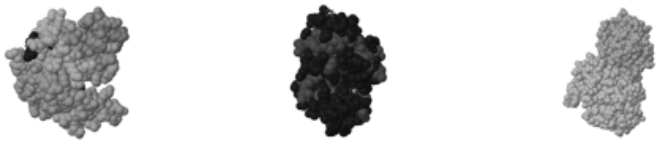
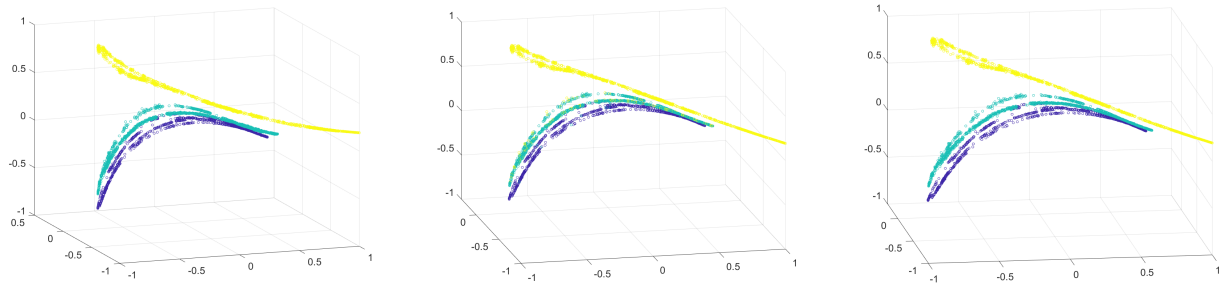
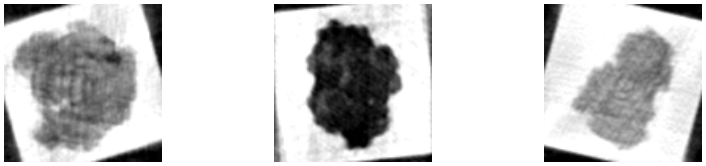


Figure 8: The original 3 conformations of Guanylate Kinase are shown in this figure.



(a) Original clusters (b) Initial Classification (c) Final Classification

Figure 9: Left: The original ground truth classification of clusters. Center: The classification provided by the moments based approach. Right: Final refined classification.



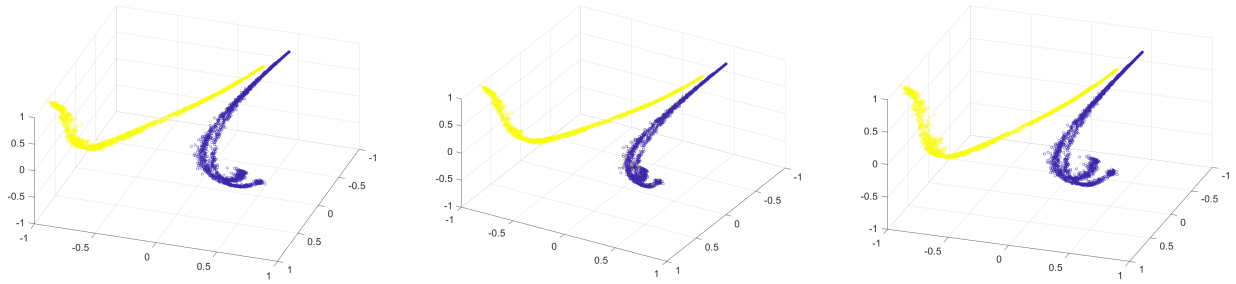
(a) 19.39% (b) 16.55% (c) 14.17%

Figure 10: The final structures estimated by our algorithm.

0.2.4 DnaK chaperone

A total of $Q = 3 \times 10^4$ projections are simulated where each projection is subjected to additive i.i.d. noise from $\mathcal{N}(0, \sigma^2)$ where $\sigma = 0.3a$.

Figure 11: The original 2 conformations of DnaK chaperone are shown in this figure.



(a) Original clusters (b) Initial Classification (c) Final Classification

Figure 12: Left: The original ground truth classification of clusters. Center: The classification provided by the moments based approach. Right: Final refined classification.

(a) 15.09% (b) 15.37%

Figure 13: The final structures estimated by our algorithm.

0.2.5 Hemoglobin

A total of $Q = 3 \times 10^4$ projections are simulated where each projection is subjected to additive i.i.d. noise from $\mathcal{N}(0, \sigma^2)$, where σ varies from $0.2a$ to $0.4a$.

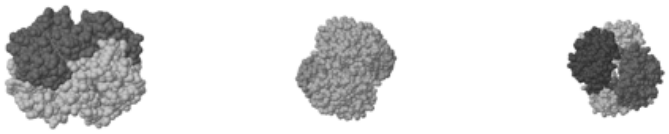
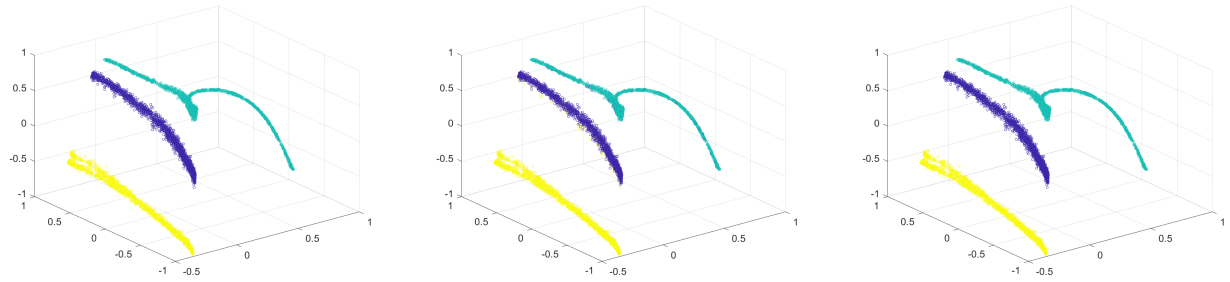
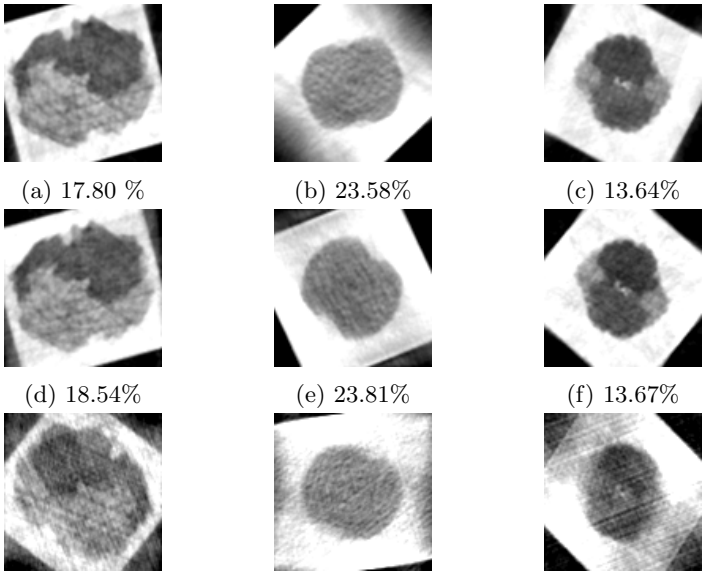


Figure 14: The original 2 conformations of Hemoglobin are shown in this figure.



(a) Original clusters (b) Initial Classification (c) Final Classification

Figure 15: Left: The original ground truth classification of clusters. Center: The classification provided by the moments based approach. Right: Final refined classification.



(a) 17.80 % (b) 23.58% (c) 13.64%

(d) 18.54% (e) 23.81% (f) 13.67%

(g) 21.26% (h) 24.39% (i) 28.55%

Figure 16: Reconstruction results in the case of different σ, along with RMSE: $\sigma = 0.2a$ (second row), $\sigma = 0.3a$ (third row) and $\sigma = 0.4a$ (fourth row).

References

[1] D. Muresan and T. Parks, "Adaptive principal components and image denoising," in *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, vol. 1, pp. 101–4, IEEE.

[2] A. Singer and H.-T. Wu, "Two-Dimensional Tomography from Noisy Projections Taken at Unknown Random Directions," *SIAM Journal on Imaging Sciences*, vol. 6, pp. 136–175, 1 2013.