DSSV 2023 Data Science, Statistics & Visualisation
EC DA 2023 European Conference on Data Analysis

Temporal Ordering and Manifold Recovery on Noisy Data

Wanjie Wang[1], Xin T. Tong[2], Yuguan Wang[3], and Yuehaw Khoo[3]

[1]National University of Singapore, Department of Statistics and Data Science, Singapore
[2]National University of Singapore, Department of Mathematics, Singapore
[3]University of Chicago, Department of Statistics, United States

The analysis of proteins and biological macromolecules is of great interest today, with the development of single-particle cryo-electron microscopy (cryo-EM). The observation Y_i follows $X_{t_i} + Noise$ because of the motions of the molecule. Since the motions repeated a hidden pattern, the ordering of t_i does not have the same ordering with i. Hence, a proper ordering of Y_i will largely improve the recovery of the functional $X(t)$.

In our work, we present a spectral method on the Laplacian matrix to order Y_i. We first reduce the noise in Y by taking the top eigenvectors of Y. Let Z be the matrix formed by these eigenvectors and we find the ordering of rows in Z. To do it, we first build the Gaussian kernel matrix on Z and then set L_Z to be the Laplacian of the kernel matrix. Ordering the second smallest eigenvector of Z will give the correct ordering of Y. We have set up the theoretical results to show consistency.