

Covariance Matrix Estimation for the Cryo-EM Heterogeneity Problem

Amit Singer, Princeton University

In cryo-electron microscopy (cryo-EM), a microscope generates a top view of a sample of randomly-oriented copies of a molecule. The cryo-EM problem is to use the resulting set of noisy 2D projection images taken at unknown directions to reconstruct the 3D structure of the molecule. In some situations, the molecule under examination exhibits structural variability, which poses a fundamental challenge in cryo-EM. The heterogeneity problem is the task of mapping the space of conformational states of a molecule. It has been previously shown that the leading eigenvectors of the covariance matrix of the 3D molecules can be used to solve the heterogeneity problem. Estimating the covariance matrix is however challenging, since only projections of the molecules are observed, but not the molecules themselves.

In this talk we show that this problem can be viewed as a noisy matrix completion problem, and we derive an estimator for the covariance matrix as a solution to a certain linear system. While we prove that the resulting estimator for the covariance matrix is consistent in the classical limit as the number of projection images grow indefinitely, an interesting open question regarding the sample complexity of the problem remains. Namely, how many images are required in order to resolve heterogeneous structures as a function of the volume size and the signal to noise ratio? We will see that solving this question requires us to extend the analysis of the spike model in principal component analysis (PCA) in high dimensions, as we encounter limiting distributions that differ from the classical Marcenko-Pastur distribution.

Joint work with G. Katsevich and A. Katsevich