

**MICHIGAN STATE UNIVERSITY**  
Department of Statistics and Probability

# COLLOQUIUM

**Kaizheng Wang**  
Columbia University

## Clustering a mixture of Gaussians with unknown covariances

**Tuesday, September 7, 2021**  
**10:20 AM - 11:10 AM 10:20 AM - 11:10 AM Eastern Time**  
**Zoom**

### **Abstract**

We investigate the clustering problem with data from a mixture of Gaussians that share a common but unknown covariance matrix. We start by considering Gaussian mixtures with two equally-sized components and derive a Max-Cut integer program based on maximum likelihood estimation. We prove its solution achieves the optimal misclassification rate when the number of samples grows linearly in the dimension, up to a logarithmic factor. For computational tractability, we develop an efficient spectral algorithm that also attains the optimal rate but requires the sample size to grow quadratically in the dimension. We conjecture that a quadratic sample size is needed for all polynomial-time algorithms, and gather numerical as well as theoretical evidence for the statistical-computational tradeoff. Finally, we generalize the Max-Cut to a k-means program that handles multi-component mixtures with possibly unequal weights. It enjoys similar optimality guarantees for mixtures of distributions that satisfy a transportation-cost inequality, encompassing Gaussian and strongly log-concave distributions.

Zoom details can be found at: <https://stt.natsci.msu.edu/stt-colloquium-zoom-info/>

*To request an interpreter or other accommodations for people with disabilities, please call the Department of Statistics and Probability at 517-355-9589.*