

Statistics Seminar  
Department of Mathematics and Statistics

<b>DATE:</b>	Thursday, February 22, 2024
<b>TIME:</b>	1:15pm - 2:15pm
<b>LOCATION:</b>	WH 100E
<b>SPEAKER:</b>	Bruce Phillips, Binghamton University
<b>TITLE:</b>	Distributed Estimation of Principal Eigenspaces.

### Abstract

Principal component analysis (PCA) is fundamental to statistical machine learning. It extracts latent principal factors that contribute to the most variation of the data. When data are stored across multiple machines, however, communication cost can prohibit the computation of PCA in a central location and distributed algorithms for PCA are thus needed. This paper proposes and studies a distributed PCA algorithm: each node machine computes the top  $K$  eigenvectors and transmits them to the central server; the central server then aggregates the information from all the node machines and conducts a PCA based on the aggregated information. We investigate the bias and variance for the resulting distributed estimator of the top  $K$  eigenvectors. In particular, we show that for distributions with symmetric innovation, the empirical top eigenspaces are unbiased, and hence the distributed PCA is “unbiased.” We derive the rate of convergence for distributed PCA estimators, which depends explicitly on the effective rank of covariance, eigengap, and the number of machines. We show that when the number of machines is not unreasonably large, the distributed PCA performs as well as the whole sample PCA, even without full access of whole data. The theoretical results are verified by an extensive simulation study. We also extend our analysis to the heterogeneous case where the population covariance matrices are different across local machines but share similar top eigenstructures.

From:

<http://www2.math.binghamton.edu/> - **Department of Mathematics and Statistics, Binghamton University**

Permanent link:

<http://www2.math.binghamton.edu/p/seminars/stat/feb222024>

Last update: **2024/02/15 19:28**

