RUTGERS–NEW BRUNSWICK
School of Arts and Sciences

Department of Statistics

# Department of Statistics Seminar Series

**Purnamrita Sarkar**
**Department of Statistics and Data Sciences**
**University of Texas at Austin**

***The Surprising Power of Simple Algorithms for Sparse PCA***

**Wednesday, October 28th, 2026**
**12:10-1:30 PM**
**Hill Center - 116**
**Light refreshments will be served in Hill Center 452 at 11:40am**

## Abstract:

Sparse principal component analysis (PCA) is a fundamental problem in high-dimensional statistics, where the goal is to estimate a leading eigenvector of a covariance matrix under a sparsity constraint. Despite its statistical appeal, sparse PCA is computationally challenging, and existing efficient algorithms often require restrictive assumptions on the covariance matrix or a good initialization. Without these restrictions, provable guarantees have largely been limited to SDP-based methods.

In this talk, I will discuss two recent results demonstrating the power of simple algorithms for sparse PCA. First, I will show that thresholding Oja's algorithm yields a single-pass, memory-efficient streaming method that requires neither a warm start nor a spiked covariance model, although its sample complexity remains suboptimal. Next, I will present a combinatorial algorithm based on truncated power iteration that achieves strong statistical guarantees for general covariance matrices, with global convergence guarantees and no need for careful initialization. Together, these results highlight both the potential and the limitations of simple, efficient algorithms for sparse PCA.

This is joint work with Syamantak Kumar, Kevin Tian, and Peiyuan Zhang.

## Bio:

Purnamrita Sarkar is an Associate Professor in the Department of Statistics and Data Sciences at the University of Texas at Austin, with a courtesy appointment in Computer Science. She received her Ph.D. from the Machine Learning Department at Carnegie Mellon University and was a postdoctoral researcher at UC Berkeley.

Dr. Sarkar's research develops the mathematical foundations of efficient and trustworthy machine learning. She designs and analyzes algorithms for high-dimensional inference and generative modeling, studying how to exploit structure in data, quantify uncertainty, and protect privacy while understanding the fundamental trade-offs between statistical accuracy and computational efficiency.