Title:
Sparse Principal Components and Subspaces: Concepts, Theory, and Computation

Speaker:
Professor Vince Vu
Department of Statistics
The Ohio State University

Time & Place:
Wednesday, October 2, 2013
Room 140 Bardeen

Cookies & Coffee @ 3:30 in Rm 1210 MSC

Abstract:
Principal components analysis (PCA) is a popular technique for unsupervised dimension reduction. Its main idea is to look for linear combinations of the variables with the largest variance. These linear combinations correspond to eigenvectors of a covariance matrix. However, in modern applications where the number of variables can be much larger than the number of samples, PCA suffers from two major weaknesses: 1) the interpretability and subsequent use of the principal directions is hindered by their dependence on all of the variables; 2) it is generally inconsistent in high-dimensions, i.e. the estimated principal directions can be noisy and unreliable. This has motivated much research over the past decade into a class of techniques called sparse PCA that combine the essence of PCA with the assumption that the phenomena of interest depend mostly on a few variables. This talk will present some recent theoretical results on sparse PCA including optimal minimax bounds for estimating the principal eigenvector and optimal minimax bounds for estimating the principal subspace spanned by the eigenvectors of a general covariance matrix. The optimal estimators turn out to be NP-hard to compute. However, it will be shown that a convex relaxation, due to d'Aspremont et al. (2007), is a near-optimal estimator of the principal eigenvector under very general conditions, and that the relaxation can be extended to $d > 1$ dimensions using a convex body called the Fantope.