



# THE UNIVERSITY OF CHICAGO

Department of Statistics

## STATISTICS COLLOQUIUM

---

PETROS DRINEAS

Computer Science Department  
Purdue University

Leverage Scores in Data Analysis

MONDAY, October 3, 2016, at 4:00 PM

Eckhart 133, 5734 S. University Avenue

*Refreshments following the seminar in Jones 111*

### ABSTRACT

The Singular Value Decomposition (SVD) of matrices and the related Principal Components Analysis (PCA) express a matrix in terms of singular vectors, which are linear combinations of all the input data and lack an intuitive physical interpretation. Motivated by the application of PCA and SVD in the analysis of populations genetics data, we will discuss the notion of leverage scores: a simple statistic that reveals columns/rows of a matrix that lie in the subspace spanned by the top principal components (left/right singular vectors). We will then use the leverage scores to present matrix decompositions that express the structure in a matrix in terms of actual columns (and/or rows) of the matrix. Such decompositions are easier to interpret in applications, since the selected columns and rows are subsets of the data. We will also discuss extensions of the leverage scores to reveal influential entries of a matrix.