

Department of Applied Mathematics and Statistics
The Johns Hopkins University

SEMINAR

S. Stanley Young
Assistant Director of Bioinformatics
National Institute of Statistical Sciences

February 1, 2007
304 Whitehead Hall
Refreshments: 3:30 p.m.
Seminar: 4:00 p.m.

NONNEGATIVE MATRIX FACTORIZATION
FOR STATISTICAL ANALYSIS

ABSTRACT

Microarrays, proteomics, metabolomics, etc., all produce data sets where there are many more predictor variables than observations: $n \ll p$. We present relatively new methods for analysis of $n \ll p$ data sets.

There are typically correlations among variables; indeed, the many variables/predictors cannot all be independent of one another. The correlations can be utilized to improve the statistical analysis. New inference methods will be presented which combine statistical testing with nonnegative matrix factorization. The methods will be demonstrated using microarray and metabolomic data sets. Papers and code for these methods can be found at <http://www.niss.org/irMF/>.

The lecture will be divided into three parts: Introduction, Robust Singular Value Decomposition, and Nonnegative Matrix Factorization.

(This is joint work with Paul Fogel, a statistical consultant living in Paris, France; and Doug Hawkins, a statistician at the University of Minnesota.)