



JOHNS HOPKINS  
BLOOMBERG  
SCHOOL of PUBLIC HEALTH

*Department of Biostatistics*

# BIOSTATISTICS SEMINAR

## Generalizing Principal Components Analysis

Genevera I. Allen

Department of Pediatrics-Neurology, Baylor College of Medicine,  
Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital,  
& Department of Statistics, Rice University, Houston, TX.

### ABSTRACT

Variables in high-dimensional data sets common in neuroimaging, metabolomics, environmetrics and time series often exhibit complex dependencies that can arise, for example, from spatial and/or temporal processes or latent network structures. Conventional multivariate analysis techniques often ignore these relationships. We propose a generalization of the singular value decomposition that is appropriate for transposable matrix data, or data in which neither the rows nor the columns can be considered independent instances. By finding the best low rank approximation of the data with respect to a transposable quadratic norm, we introduce a novel decomposition, entitled the Generalized least squares Matrix Decomposition (GMD). This decomposition can be used to perform generalized principal components analysis (GPCA) which directly accounts for dependencies in the data. We also introduce a framework for regularizing the factors of the GMD that can be used to perform sparse GPCA, functional GPCA, non-negative GPCA, and sparse non-negative GPCA. Additionally, we develop fast numerical algorithms allowing one to apply our methods to massive data sets. Through simulations and real data examples on functional MRI data and NMR spectroscopy, we demonstrate the utility of GPCA and sparse GPCA for dimension reduction, signal recovery, and feature selection with high-dimensional transposable data.

Keywords: matrix decomposition, singular value decomposition, transposable data, principal components analysis, sparse principal components analysis.

**The Johns Hopkins Bloomberg School of Public Health  
Department of Biostatistics, Wednesday, October 5, 2011  
Room W2030 School of Public Health, 4:00-5:00pm (Refreshments: 3:30)**

For disability access information or listening devices, please contact the [Office of Support Services](#) at 410-955-1197 or on the Web at [www.jhsph.edu/SupportServices](http://www.jhsph.edu/SupportServices). EO/AA