General Consistency Results of PCA in High Dimension

Prof. Dr. Sungkyu Jung
(University of Pittsburgh)

June 29, 2012

Principal component Analysis is a widely used method for dimensionality reduction and visualization of multidimensional data. It becomes common in modern data analytic situation that the dimension d of the observation is much larger than the sample size n. This leads to a new domain in asymptotic studies of the estimated principal component analysis, that is, in terms of the limit of d. A unified framework for assessing the consistency of principal component estimates in a wide range of asymptotic settings is provided. In particular, our result works for any ratio of dimension and sample size, $d/n \to c$, $c \in [0, \infty]$. We apply this framework to two different statistical situations. When applied to a factor model, we obtain a unified view on the sufficient condition for the consistency of principal component analysis. We propose to use time-varying principal components to model multivariate longitudinal data with an irregular grid. A sufficient condition for the consistency of the estimates is obtained by the proposed tool. Simulation results and real data analysis are included. (This is joint work with Jason Fine and J. S. Marron)