Jung, Sungkyu; Marron, J. S.
PCA consistency in high dimension, low sample size context. (English) Zbl 1191.62108

Summary: Principal Component Analysis (PCA) is an important tool of dimension reduction especially when the dimension (or the number of variables) is very high. Asymptotic studies where the sample size is fixed, and the dimension grows [i.e., High Dimension, Low Sample Size (HDLSS)] are becoming increasingly relevant. We investigate the asymptotic behavior of Principal Component (PC) directions. HDLSS asymptotics are used to study consistency, strong inconsistency and subspace consistency. We show that if the first few eigenvalues of a population covariance matrix are large enough compared to the others, then the corresponding estimated PC directions are consistent or converge to the appropriate subspace (subspace consistency) and most other PC directions are strongly inconsistent. Broad sets of sufficient conditions for each of these cases are specified and the main theorem gives a catalogue of possible combinations. In preparation for these results, we show that the geometric representation of HDLSS data holds under general conditions, which includes a $\rho$-mixing condition and a broad range of sphericity measures of the covariance matrix.

MSC:
62H25 Factor analysis and principal components; correspondence analysis
34L20 Asymptotic distribution of eigenvalues, asymptotic theory of eigenfunctions for ordinary differential operators
62F12 Asymptotic properties of parametric estimators
15A18 Eigenvalues, singular values, and eigenvectors

Keywords:
principal component analysis; sample covariance matrix; $\rho$-mixing; high dimension; low sample size data; nonstandard asymptotics; consistency and strong inconsistency; spiked population model

Full Text: DOI arXiv

References:


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.