High-dimensional inference on a cross data matrix-based method

Shao-Hsuan Wang (National Central University)

Abstract

The concept of the cross-data matrix originates from the work of Yata and Aoshima (2010), who demonstrated that the cross-data matrix-based principal component analysis (CDM-PCA) method can effectively reduce noise and enhance the performance of principal component analysis (PCA) in high-dimensional, low-sample-size settings. This innovative approach has inspired numerous subsequent studies. For instance, Wang, Huang, and Chen (2020) established the asymptotic normality of estimates for principal component directions, while Wang and Huang (2022) derived finite-sample approximations and explored the asymptotic behavior of CDM-based PCA through matrix perturbation theory. More recently, Hung and Huang (2023) introduced a more stable variant of CDM-PCA, termed product-PCA (PPCA). This formulation offers a more convenient structure for theoretical analysis and has been shown to be more robust than PCA in preserving the correct ordering of leading eigenvalues, even in the presence of outliers.

In this talk, I will discuss recent advances in the cross-data matrix-based methods for high-dimensional data analysis, which will be presented in two parts. First, I will introduce cross-data matrix-based Multilinear Principal Component Analysis (CDM-MPCA) along with its numerical studies. In the second part, I will present the limiting spectral distribution (LSD) for the singular values of large cross-data matrix-based sample covariance matrix. Additionally, I will compare this distribution with the Marchenko–Pastur law (MP law), which characterizes the asymptotic behavior of the singular values of large sample covariance matrix.