

Augmented James–Stein estimation for leading eigenvectors and eigenspaces in high dimensions

Giheon Seong¹, Seungki Hong², Sungkyu Jung³

¹Department of Statistics, Seoul National University, heon1998@snu.ac.kr

²Meritz Fire & Marine Insurance Co., Ltd., seungki.hong@meritz.co.kr

³Department of Statistics and Institute for Data Innovation in Science, Seoul National University, sungkyu@snu.ac.kr

Abstract

Building on the James–Stein approach to leading eigenvector estimation (Goldberg and Kercheval, 2023), we develop a data-adaptive augmented James–Stein shrinkage framework for estimating leading eigenvectors and eigenspaces under a generalized spiked population model, in the high-dimensional regime where the dimension p and sample size n grow proportionally. For each spiked eigenvector, we construct an augmented target subspace that combines auxiliary information, either from domain knowledge or prior information, with the remaining sample spiked eigenvectors. This augmentation allows information shared across the sample spiked components to be exploited while retaining a fully data-driven shrinkage rule. We show that the resulting eigenvector estimator strictly improves upon standard PCA whenever the target subspace contains nonvanishing information about the population eigenvector, while asymptotically reverting to PCA when the target is uninformative. The individual estimators further yield a nested sequence of estimators for all leading spiked eigenspaces, with analogous dominance properties. For the full spiked eigenspace, the proposed estimator also strictly improves upon the existing HDLSS-motivated shrinkage estimator in the proportional high-dimensional regime, while becoming asymptotically equivalent to it under HDLSS asymptotics. Simulation studies demonstrate substantial finite-sample gains and robustness to misspecification of the number of spikes.