

## A Data-Adaptive Principal Component Analysis

Yaeji Lim\*

Seoul National University, Seoul, South Korea [yaeji@snu.ac.kr](mailto:yaeji@snu.ac.kr)

Hee-Seok Oh

Seoul National University, Seoul, South Korea [heeseok.oh@gmail.com](mailto:heeseok.oh@gmail.com)

This paper studies a data-adaptive principal component analysis (PCA) that does not require prior information of data distribution. The ordinary PCA is useful for dimension reduction and for identifying important features of data that consist of a large number of interrelated variables. However, it is stringent to the Gaussian assumption of the data, and therefore may not be efficient for analyzing real observations that may be non-Gaussian distributed, such as skewed or heavy-tailed data. To extend the scope of PCA to non-Gaussian distributed data, a new approach for PCA is proposed. The core of the methodology is the use of a composite quantile, which is a weighted linear combination of convex loss functions instead of the square loss function, and the weights are determined data-adaptively. Furthermore, a practical algorithm to implement the data-adaptive PCA is derived. Results from a simulation study and real data examples demonstrate the promising empirical properties of the proposed approach.

Key Words: High-dimensional data, Pseudo data, Quantile estimation, Robustness