Accepted Manuscript

Projection tests for high-dimensional spiked covariance matrices

Wenwen Guo, Hengjian Cui

PII: S0047-259X(17)30755-8
DOI: https://doi.org/10.1016/j.jmva.2018.08.009
Reference: YJMVA 4404

To appear in: *Journal of Multivariate Analysis*

Received date : 7 December 2017

Please cite this article as: W. Guo, H. Cui, Projection tests for high-dimensional spiked covariance matrices, *Journal of Multivariate Analysis* (2018), https://doi.org/10.1016/j.jmva.2018.08.009

Projection tests for high-dimensional spiked covariance matrices

Wenwen Guo[a,*], Hengjian Cui[1]

[a]*Capital Normal University, Beijing*

Abstract

Testing the existence of low-dimensional perturbations or signals is very important, e.g., in factor analysis and signal processing. This paper aims to develop new tests for high-dimensional spiked covariance matrices based on a projection approach. The asymptotic distribution of the proposed tests is obtained under regularity conditions. We further explore a power enhancement technique under covariance matrix sparsity. The finite-sample enhanced power performance of the proposed tests is shown through simulations. A microarray dataset is used for illustration purposes.

Keywords: Large p small n, Power enhancement technique, Projection, Spiked covariance matrix

1. Introduction

Considerable attention has been paid to testing the identity and sphericity of covariance matrices. In modern applications, the matrix is often large, with the number p of variables comparable to, or much larger than, the sample size n. In order to accommodate high dimensionality, new tests based on entropy or quadratic loss have been proposed. Under entropy loss, Bai et al. [2] proposed a modified likelihood ratio statistic for testing identity when $p/n \to c \in (0, 1)$. Further work about the likelihood ratio test in high-dimensional settings is reported in [13]. Zhang et al. [26] proposed empirical likelihood ratio procedures for testing whether a covariance matrix equals a given one or has a banded structure. Under quadratic loss, Ledoit and Wolf [17] built a test statistic by measuring the Euclidean distance between the sample covariance matrix and the null matrix, and they obtained asymptotic properties in a Gaussian framework when $p/n \to c < \infty$.

To avoid estimating the covariance matrix when $p > n$, Chen et al. [10] investigated testing procedures based on U-statistics for both the identity and the sphericity hypothesis. Li and Chen [18] proposed tests for equality of covariance matrices applicable for "large p, small n" situations. Cai and Ma [8] considered testing the equality of a covariance matrix to a given one; they characterized the boundary separating the testable region from the non-testable region when p/n is bounded. Recently, Peng et al. [24] considered improving the power of testing for identity and sphericity by employing banding estimators for high-dimensional covariance matrices.

Recent studies place an emphasis on the existence of low-dimensional perturbations or signals in the data, which is a special alternative to the identity or sphericity of the covariance matrix. To be exact, the corresponding alternative covariance matrix is proportional to the sum of the identity matrix and a matrix of finite rank K. Johnstone [14] calls it a "spiked covariance." The practical justification for analyzing such matrices comes from various sources, such as signal processing and factor analysis; see [4, 5, 15, 16, 21–23] for a few examples.

In a high-dimensional setting, a sparsity assumption was proposed to ensure the asymptotic properties of the statistics. Cai et al. [9] considered both minimax and adaptive estimation of the principal subspace in sparse principal component analysis and established the optimal rates of convergence for estimation of the principal subspace. Berthet and Rigollet [4] investigated the detection of sparse principal component problems, and they proposed a test based on a conservative critical value in order to control the type I error of the test. However, to the best of our knowledge, there has been no attempt to address the problem of testing spiked covariance matrices, which motivates this work.

*Corresponding author

Email addresses: wwguo@cnu.edu.cn (Wenwen Guo), hjcui@bnu.edu.cn (Hengjian Cui)