Detection of Signal in the Spiked Rectangular Models

Ji Hyung Jung [1] **Hye Won Chung** [2] **Ji Oon Lee** [1 3]

Abstract

We consider the problem of detecting signals in the rank-one signal-plus-noise data matrix models that generalize the spiked Wishart matrices. We show that the principal component analysis can be improved by pre-transforming the matrix entries if the noise is non-Gaussian. As an intermediate step, we prove a sharp phase transition of the largest eigenvalues of spiked rectangular matrices, which extends the Baik–Ben Arous–Péché (BBP) transition. We also propose a hypothesis test to detect the presence of signal with low computational complexity, based on the linear spectral statistics, which minimizes the sum of the Type-I and Type-II errors when the noise is Gaussian.

1. Introduction

Detecting a low-rank structure or signal in a high-dimensional noisy data is one of the most fundamental problems in statistics and data science (Johnstone, 2001; Onatski et al., 2013; 2014; Abbe, 2017). In the case where the data is a matrix and the signal is a vector, it is natural to consider spiked random matrices, which includes the spiked Wigner matrix and the spiked Wishart matrix. In these models, the signal is in the form of rank-1 mean matrix (spiked Wigner matrix) or rank-1 perturbation of the identity in the covariance matrix (spiked Wishart matrix). In this paper, we consider the following rectangular random matrix models that generalize the spiked Wishart matrix:

- Rectangular matrix *with spiked mean* (additive model): the data matrix is of the form

  $$\sqrt{\lambda}\boldsymbol{u}\boldsymbol{v}^T + X,$$

  where X is an $M \times N$ random i.i.d. matrix whose entries are centered with variance N^{-1}, $\boldsymbol{u} \in \mathbb{R}^M$, $\boldsymbol{v} \in \mathbb{R}^N$ with $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1$. The parameter λ corresponds to the signal-to-noise ratio (SNR).

- Rectangular matrix *with spiked covariance* (multiplicative model): the data matrix is of the form

  $$(I + \lambda\boldsymbol{u}\boldsymbol{u}^T)^{1/2}X,$$

  where X is an $M \times N$ random i.i.d. matrix whose entries are centered with variance N^{-1}, $\boldsymbol{u} \in \mathbb{R}^M$ with $\|\boldsymbol{u}\| = 1$. The parameter λ corresponds to the SNR.

Note that in the rectangular matrix with spiked covariance, and also in the rectangular matrix with spiked mean under an additional assumption that the entries of $\boldsymbol{u}$ and $\boldsymbol{v}$ are centered, the population covariance is

$$\Sigma = I + \lambda\boldsymbol{u}\boldsymbol{u}^T.$$

In the special case where the entries of $\boldsymbol{v}$ are i.i.d. Gaussians, the two models coincide.

If SNR λ is sufficiently large, we can easily detect (and recover) the signal by methods such as principal component analysis (PCA). Even under the high-dimensional assumption $M, N \to \infty$ with $M/N \to d_0 \in (0, \infty)$, the signal can be reliably detected by PCA if $\lambda > \sqrt{d_0}$. (For the use of PCA in the high-dimensional setting, we refer to (Johnstone, 2007).) On the other hand, if $\lambda \in (0, \sqrt{d_0})$, the distribution of the largest eigenvalue coincides with that of the null model $\lambda = 0$. This sharp transition in the behavior of the largest eigenvalue is known as the BBP transition after the seminal work by Baik, Ben Arous, and Péché (Baik et al., 2005). (See Section 2.2.)

On the other hand, in the subcritical case $\lambda < \sqrt{d_0}$, if the noise X is Gaussian and the signal $\boldsymbol{u}$ (and also $\boldsymbol{v}$ for a rectangular matrix with spiked mean) is drawn uniformly from the unit sphere, known as the spherical prior, then no test can reliably detect the signal. (See Section 2.3.) Thus, it is natural to ask the following questions:

- Is the threshold for reliable detection (i.e., with probability $1 - o(1)$ as $M, N \to \infty$) lower than $\sqrt{d_0}$ if the noise is non-Gaussian?

- Can we design an efficient algorithm to weakly detect the signal (i.e., better than a random guess) for the subcritical case?

[1]Department of Mathematical Sciences, KAIST, Daejeon, Korea [2]School of Electrical Engineering, KAIST, Daejeon, Korea [3]School of Mathematics, KIAS, Seoul, Korea. Correspondence to: Hye Won Chung <hwchung@kaist.ac.kr>, Ji Oon Lee <jioon.lee@kaist.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021. Copyright 2021 by the author(s).

We aim to answer these questions in this paper.

1.1. Main contributions

Our main contributions are as follows:

- We prove that the PCA can be improved by an entry-wise transformation if the noise is non-Gaussian, under a mild assumption on the distribution (prior) of the spike.
- We propose a universal test to detect the presence of signal with low computational complexity, based on the linear spectral statistics (LSS). The test does not require any prior information on the signal, and if the noise is Gaussian the error of the proposed test is optimal.

Heuristically, the SNR can be increased through an entry-wise transformation and it can be easily seen for a rectangular matrix (additive model) of the form $Y = \sqrt{\lambda}\boldsymbol{u}\boldsymbol{v}^T + X$. If $|u_i v_j| \ll X_{ij}$, then by applying a function q entrywise to $\sqrt{N}Y$, we obtain a transformed matrix whose entries are

$$\begin{aligned} q(\sqrt{N}Y_{ij}) &= q(\sqrt{N}X_{ij} + \sqrt{\lambda N}u_i v_j) \\ &\approx q(\sqrt{N}X_{ij}) + \sqrt{\lambda N}q'(\sqrt{N}X_{ij})u_i v_j, \end{aligned}$$

where the approximation is due to the Taylor expansion. It can be shown that the coefficient $q'(\sqrt{N}X_{ij})$ in the second term in the right side can be replaced by its expectation with negligible error. (See Section B.2 of Supplementary Material for the proof.) Thus,

$$\begin{aligned} q(\sqrt{N}Y_{ij}) &= q(\sqrt{N}X_{ij} + \sqrt{\lambda N}u_i v_j) \\ &\approx \sqrt{N}\left(\frac{q(\sqrt{N}X_{ij})}{\sqrt{N}} + \sqrt{\lambda}\mathbb{E}[q'(\sqrt{N}X_{ij})]u_i v_j\right), \end{aligned}$$

and the transformed matrix is of the form $\sqrt{\lambda'}\boldsymbol{u}\boldsymbol{v}^T + Q$ after normalization, which yields another spiked rectangular matrix with different SNR. By optimizing the SNR of the transformed matrix, we find that the SNR is effectively increased (or equivalently, the threshold $\sqrt{d_0}$ is lowered) in the PCA for the transformed matrix. The change of the SNR can be rigorously proved; see Theorem 5 for a precise statement. We remark a similar idea was also discussed in (Montanari et al., 2018) without rigorous proof.

The corresponding result is not known, to our best knowledge, for the multiplicative model of the form $Y = (I + \lambda\boldsymbol{u}\boldsymbol{u}^T)^{1/2}X =: (I + \gamma\boldsymbol{u}\boldsymbol{u}^T)X$. (Here, $\lambda = 2\gamma + \gamma^2$.) The analysis is significantly more involved in this case due to the following reason: When applying a function q entrywise to $\sqrt{N}Y$, we find that

$$\begin{aligned} q(\sqrt{N}Y_{ij}) &= q\Big(\sqrt{N}X_{ij} + \gamma\sqrt{N}\sum_k u_i u_k X_{kj}\Big) \\ &\approx q(\sqrt{N}X_{ij}) + \gamma\sqrt{N}q'(\sqrt{N}X_{ij})\sum_k u_i u_k X_{kj} \\ &\approx \sqrt{N}\left(\frac{q(\sqrt{N}X_{ij})}{\sqrt{N}} + \gamma\mathbb{E}[q'(\sqrt{N}X_{ij})]\sum_k u_i u_k X_{kj}\right), \end{aligned}$$

and the transformed matrix is of the form $\gamma'\boldsymbol{u}\boldsymbol{u}^T X + Q$, which is not a spiked rectangular matrix anymore. (Note that Q depends on X and thus it cannot be considered as an additive model, either.)

In Theorem 6 in Section 3.1, we prove the effective change of the SNR for the multiplicative model. The proof of Theorem 6 is based on a generalized version of the BBP transition that works with the matrix of the form $\gamma\boldsymbol{u}\boldsymbol{u}^T X + Q$. Applying various results and techniques from random matrix theory, we introduce a general strategy to prove a BBP-type transition and apply it to the transformed matrix.

It is notable that the optimal entrywise transform is different from the one for the additive model. For the additive model, the optimal transform is given by $-g'/g$, where g is the density function of the noise entry. However, for the multiplicative model, the optimal transform is a linear combination of the function $-g'/g$ and the identity mapping. Heuristically, it is due to that the effective SNR depends not only on γ' but also on the correlation between X and Q; the former is maximized when the transform is $-g'/g$ while the latter is maximized when the transform is the identity mapping. We also remark that the effective SNR after the optimal entrywise transform is larger in the additive model, which suggests that the detection problem is fundamentally harder for the multiplicative model.

When it is impossible to reliably detect the signal, the next goal is the weak detection, which is basically the hypothesis testing problem between the null model and the alternative model that the spike exists in the data. As predicted by the Neyman–Pearson lemma, the likelihood ratio (LR) test is optimal in the sense that it minimizes the sum of the Type-I error and the Type-II error. The limit of the log-LR was proved to be Gaussian for both the additive model and the multiplicative model with Gaussian noise (Onatski et al., 2013; El Alaoui & Jordan, 2018) from which the limiting optimal error can be readily deduced.

However, LR tests require substantial information on the prior, which is not available in many applications. Following the idea in (Chung & Lee, 2019), we propose a test based on the LSS, which does not require any knowledge on the spike or the noise. We prove in Corollary 9 (see also Remark 10) that the error of the proposed test is optimal if the noise is Gaussian.

The proposed test is applicable even when the noise is non-Gaussian. It is expected that the weak detection based on the proposed test will perform better after the entrywise transform, which was proved for spiked Wigner models (Chung & Lee, 2019). This will be discussed in a future paper. We also conjecture that with the entrywise transform our test will be optimal when the noise is non-Gaussian, but it is beyond our scope as the optimal error of the weak detection for non-Gaussian noise is not known, even for spiked Wigner models.

1.2. Related works

Spiked rectangular model was introduced by Johnstone (Johnstone, 2001). The transition of the largest eigenvalue was proved by Baik, Ben Arous, and Péché (Baik et al., 2005) for spiked complex Wishart matrices and generalized by Benaych-Georges and Nadakuditi (Benaych-Georges & Nadakuditi, 2011; 2012). For more results from random matrix theory about the largest eigenvalue and the corresponding eigenvector of a spiked rectangular matrix, we refer to (Bloemendal et al., 2016) and references therein.

The testing problem for spiked Wishart matrices with the spherical prior and Gaussian noise was considered by Onatski, Moreira, and Hallin (Onatski et al., 2013; 2014), where they proved the optimal error of the hypothesis test. It is later extended to the case where the entries of the spikes are i.i.d. with bounded support (i.i.d. prior) by El Alaoui and Jordan (El Alaoui & Jordan, 2018). LSS-based tests in the spiked rectangular models with the standardized entries were considered by Dobriban (Dobriban, 2017).

The improved PCA based on the entrywise transformation was considered for spiked Wigner models in (Lesieur et al., 2015; Perry et al., 2018), where the transformation is chosen to maximize the effective SNR of the transformed matrix. Detection problems for spiked Wigner models were also considered, where the analysis is typically easier due to its symmetry and canonical connection with spin glass models. For more results on the spiked Wigner models, we refer to (Montanari et al., 2017; Perry et al., 2018; El Alaoui et al., 2020; Chung & Lee, 2019) and references therein.

1.3. Organization of the paper

The rest of the paper is organized as follows. In Section 2, we precisely define the model and introduce previous results. In Section 3, we state our results on the improved PCA and illustrate the improvement of PCA by numerical experiments. In Section 4, we state our results on the hypothesis testing and the central limit theorems for the linear spectral statistics. We conclude the paper in Section 5 with the summary of our results and future research directions. Details of the numerical simulations and the proofs of the technical results can be found in Supplementary Material.

2. Preliminaries

2.1. Definition of the model

We begin by precisely defining the model we consider in this paper. The noise matrix has the following properties.

Definition 1 (Rectangular matrix)**.** We say an $M \times N$ random matrix $X = (X_{ij})$ is a (real) rectangular matrix if X_{ij} ($1 \le i \le M$, $1 \le j \le N$) are independent real random variables satisfying the following conditions:

- For all i, j, $\mathbb{E}[X_{ij}] = 0$, $N\mathbb{E}[X_{ij}^2] = 1$, $N^{\frac{3}{2}}\mathbb{E}[X_{ij}^3] = w_3$, and $N^2\mathbb{E}[X_{ij}^4] = w_4$ for some constants w_3, w_4.
- For any positive integer p, there exists C_p, independent of N, such that $N^{\frac{p}{2}}\mathbb{E}[X_{ij}^p] \le C_p$ for all i, j.

The spiked rectangular matrices are defined as follows.

Definition 2 (Spiked rectangular matrix - additive model)**.** We say an $M \times N$ random matrix $Y = \sqrt{\lambda}\boldsymbol{u}\boldsymbol{v}^T + X$ is a rectangular matrix with spiked mean $\boldsymbol{u}$, $\boldsymbol{v}$ and SNR λ if $\boldsymbol{u} = (u_1, u_2, \dots, u_M)^T \in \mathbb{R}^M$, $\boldsymbol{v} = (v_1, v_2, \dots, v_N)^T \in \mathbb{R}^N$ with $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1$, and X is a rectangular matrix.

Definition 3 (Spiked rectangular matrix - multiplicative model)**.** We say an $M \times N$ random matrix $Y = (I + \lambda\boldsymbol{u}\boldsymbol{u}^T)^{1/2}X$ is a rectangular matrix with spiked covariance $\boldsymbol{u}$ and SNR λ if $\boldsymbol{u} = (u_1, u_2, \dots, u_M)^T \in \mathbb{R}^M$ with $\|\boldsymbol{u}\| = 1$ and X is a rectangular matrix.

We assume throughout the paper that $\lambda \ge 0$ and $\frac{M}{N} \to d_0 \in (0, \infty)$ as $M, N \to \infty$.

2.2. Principal component analysis

Let $S = YY^T$ be the sample covariance matrix (Gram matrix) derived from a spiked rectangular matrix Y. The empirical spectral measure of S converges to the Marchenko–Pastur law μ_{MP}, i.e., if we denote by $\mu_1 \ge \mu_2 \ge \dots \mu_M$ the eigenvalues of S, then

$$\frac{1}{M}\sum_{i=1}^{M} \delta_{\mu_i}(x)\mathrm{d}x \to \mathrm{d}\mu_{MP}(x) \tag{1}$$

weakly in probability as $M, N \to \infty$, where for $M \le N$

$$\mathrm{d}\mu_{MP}(x) = \frac{\sqrt{(x - d_-)(d_+ - x)}}{2\pi d_0 x}\mathbf{1}_{(d_-, d_+)}(x)\mathrm{d}x, \tag{2}$$

with $d_\pm = (1 \pm \sqrt{d_0})^2$. The largest eigenvalue has the following (almost sure) limit:

- If $\lambda > \sqrt{d_0}$, then $\mu_1 \to (1 + \lambda)(1 + \frac{d_0}{\lambda})$.
- If $\lambda < \sqrt{d_0}$, then $\mu_1 \to d_+ = (1 + \sqrt{d_0})^2$.

This in particular shows that the detection can be reliably done by PCA if $\lambda > \sqrt{d_0}$.

2.3. Likelihood ratio

Denote by $\mathbb{P}_1$ the joint probability of the data Y, a spiked rectangular matrix, with $\lambda = \omega > 0$ and $\mathbb{P}_0$ with $\lambda = 0$. When the noise is Gaussian, the likelihood ratio $\mathcal{L}(Y;\lambda)$ of $\mathbb{P}_1$ with respect to $\mathbb{P}_0$ is given by

$$\int \det(I + \omega \boldsymbol{u}\boldsymbol{u}^T)^{-\frac{N}{2}} \times \exp\left(\frac{N\lambda}{2(1+\omega\|\boldsymbol{u}\|^2)} \sum_{i=1}^{M}\sum_{j=1}^{N} (YY^T)_{ij} u_i u_j \right) \mathrm{d}\mathcal{P}_{\boldsymbol{u}}$$

for the multiplicative model (Definition 3) and

$$\int \exp\left(N \sum_{i=1}^{M}\sum_{j=1}^{N} \left[\sqrt{\omega} Y_{ij} u_i v_j - \frac{\omega}{2} u_i^2 v_j^2 \right] \right) \mathrm{d}\mathcal{P}_{\boldsymbol{u}} \mathrm{d}\mathcal{P}_{\boldsymbol{v}}$$

for the additive model (Definition 2). Here, $\mathcal{P}_{\boldsymbol{u}}$ and $\mathcal{P}_{\boldsymbol{v}}$ are the prior distributions of $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively.

If $\omega < \sqrt{d_0}$, for both models with the spherical prior where the spike is drawn uniformly from the unit sphere, the log-LR has the Gaussian limit; as $N \to \infty$, it converges to

$$\mathcal{N}\left(\frac{1}{4}\log\left(1 - \frac{\omega^2}{d_0}\right), -\frac{1}{2}\log\left(1 - \frac{\omega^2}{d_0}\right) \right)$$

under the null hypothesis $\boldsymbol{H}_0 : Y \sim \mathbb{P}_0$ and

$$\mathcal{N}\left(-\frac{1}{4}\log\left(1 - \frac{\omega^2}{d_0}\right), -\frac{1}{2}\log\left(1 - \frac{\omega^2}{d_0}\right) \right)$$

under the alternative hypothesis $\boldsymbol{H}_1 : Y \sim \mathbb{P}_1$. The same result also holds for the additive model with Rademacher prior. The sum of the Type-I error and the Type-II error of the likelihood ratio test

$$\begin{aligned} \mathrm{err}(\omega) &:= \mathbb{P}(L(Y;\omega) > 1 | \boldsymbol{H}_0) + \mathbb{P}(L(Y;\omega) \le 1 | \boldsymbol{H}_1) \\ &\to \mathrm{erfc}\left(\frac{1}{4} \sqrt{-\log\left(1 - \frac{\omega^2}{d_0}\right)} \right) \end{aligned} \tag{3}$$

as $N \to \infty$. We remark that it is the minimal error among all tests as Neyman–Pearson lemma asserts. This in particular shows that the reliable detection of signal is impossible with Gaussian noise when $\omega < \sqrt{d_0}$.

2.4. Linear spectral statistics

The proof of the Gaussian convergence of the LR in (Baik & Lee, 2016; 2020) is based on the recent study of linear spectral statistics, defined as

$$L_Y(f) = \sum_{i=1}^{M} f(\mu_i) \tag{4}$$

for a function f, where $\mu_1 \ge \mu_2 \ge \dots \mu_M$ are the eigenvalues of $S = YY^T$. As Marchenko–Pastur law in (1) suggests, it is required to consider the fluctuation of the LSS about

$$M \int_{d_-}^{d_+} f(x) \, \mathrm{d}\mu_{MP}(x).$$

The CLT for the LSS is the statement

$$\begin{aligned} &\left(L_Y(f) - M \int_{d_-}^{d_+} f(x) \, \mathrm{d}\mu_{MP}(x) \right) \\ &\Rightarrow \mathcal{N}(m_Y(f), V_Y(f)), \end{aligned} \tag{5}$$

where the right-hand side is the Gaussian random variable with the mean $m_Y(f)$ and the variance $V_Y(f)$. The CLT was proved for the null case ($\lambda = 0$). We will show that the CLT also holds under the alternative and the mean $m_Y(f)$ depends on λ while the variance $V_Y(f)$ does not.

3. Main Result I - Improved PCA

In this section, we state our first main results on the improvement of PCA by entrywise transformations and provide the results from numerical experiments.

3.1. Improved PCA

Let $\mathcal{P}$ be the distribution of the normalized entry $\sqrt{N}X_{ij}$ whose density function is g. As we discussed in Section 1.1, applying a function q to the additive model in Definition 3 approximately yields another rectangular matrix

$$\frac{q(\sqrt{N}X_{ij})}{\sqrt{N}} + \sqrt{\lambda}\mathbb{E}[q'(\sqrt{N}X_{ij})] u_i v_j. \tag{6}$$

Suppose that $q(\sqrt{N}X_{ij}) = \sqrt{N}Q_{ij}$ is with mean 0 and variance 1. Then, the effective SNR of the transformed matrix is $\lambda(\mathbb{E}[q'(\sqrt{N}X_{ij})])^2$, which is maximized when $q(x)$ is a multiple of $-g'(x)/g(x)$.

For the multiplicative model in Definition 3, applying a function q approximately yields a transformed matrix of the form $Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X$ as discussed in Section 1.1, where we set $\widehat{\gamma} = \gamma\mathbb{E}[q'(\sqrt{N}X_{ij})]$. The sample covariance matrix generated by it is

$$\begin{aligned} &(Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X)(Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X)^T \\ &= QQ^T + \widehat{\gamma} Q X^T \boldsymbol{u}\boldsymbol{u}^T + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X Q^T \\ &\quad + \widehat{\gamma}^2 \boldsymbol{u}\boldsymbol{u}^T X X^T \boldsymbol{u}\boldsymbol{u}^T. \end{aligned}$$

Conditioning on $\boldsymbol{u}$, its expectation is $(I + \lambda_{SNR}\boldsymbol{u}\boldsymbol{u}^T)$, where the effective SNR λ_{SNR} is

$$\begin{aligned} &2\widehat{\gamma}\mathbb{E}[\sqrt{N}X_{ij} q(\sqrt{N}X_{ij})] + \widehat{\gamma}^2 \\ &= 2\gamma\mathbb{E}[q'(\sqrt{N}X_{ij})]\mathbb{E}[\sqrt{N}X_{ij} q(\sqrt{N}X_{ij})] \\ &\quad + \gamma^2(\mathbb{E}[q'(\sqrt{N}X_{ij})])^2 \end{aligned}$$

We can find that λ_{SNR} is maximized when $q(x)$ is a multiple of $-g'(x)/g(x)+\alpha x$ for some constant α.

In this section, we rigorously prove our heuristic argument and show the detection threshold of PCA can be lowered by applying the entrywise transformations above. We introduce the following assumptions for the spike and the noise.

Assumption 4. For the spike $\boldsymbol{u}$ (and also $\boldsymbol{v}$ in the additive model), we assume either

1. the spherical prior, i.e., $\boldsymbol{u}$ (and $\boldsymbol{v}$) are drawn uniformly from the unit sphere, or
2. the i.i.d. prior, i.e., the entries $u_1, u_2, \dots, u_M$ (respectively, $v_1, v_2, \dots, v_N$) are i.i.d. random variables with mean zero and variance M^{-1} (respectively N^{-1}) such that for any integer $p > 2$

$$\mathbb{E}|u_i|^p, \mathbb{E}|v_j|^p \le \frac{C_p}{M^{1+(p-2)\phi}}$$

for some (N-independent) constants $C_p > 0$ and $\phi > \frac{1}{4}$, uniformly on i and j.

For the noise, let $\mathcal{P}$ be the distribution of the normalized entries $\sqrt{N}X_{ij}$. We assume the following:

1. The density function g of $\mathcal{P}$ is smooth, positive everywhere, and symmetric (about 0).
2. For any fixed D, the D-th moment of $\mathcal{P}$ is finite.
3. The function $h = -g'/g$ and its all derivatives are polynomially bounded in the sense that $|h^{(\ell)}(w)| \le C_\ell |w|^{C_\ell}$ for some constant C_ℓ depending only on ℓ.

Note that the signal is not necessarily delocalized, i.e., some entries of the signal can be much larger than $N^{-1/2}$.

We remark that some conditions in Assumption 4, especially the i.i.d. prior and the finiteness of all moments of $\mathcal{P}$, are technical constraints and our results hold under weaker assumptions. We also remark that if $\sqrt{M}u_i$ (and $\sqrt{N}v_j$) are i.i.d. random variables, independent of M (and N), whose all moments are finite, Assumption 4 is satisfied with $\phi = \frac{1}{2}$.

Given the data matrix Y, we consider a family of the entrywise transformations of the form $h_\alpha(x) = -g'(x)/g(x) + \alpha x$ and transformed matrices $\widetilde{Y}^{(\alpha)}$ whose entries are

$$\widetilde{Y}_{ij}^{(\alpha)} = \frac{1}{\sqrt{(\alpha^2+2\alpha+F_g)N}} h_\alpha(\sqrt{N}Y_{ij}), \qquad (7)$$

where the Fisher information F_g of g is given by

$$F_g = \int_{-\infty}^{\infty} \frac{(g'(x))^2}{g(x)} \mathrm{d}x.$$

Note that $F_g \ge 1$ where the equality holds only if g is the standard Gaussian.

For the additive model, we show that the effective SNR of the transformed matrix for PCA is λF_g.

Theorem 5. *Let Y be a spiked rectangular matrix in Definition 2 that satisfy Assumption 4. Let $\widetilde{Y} \equiv \widetilde{Y}^{(0)}$ be the transformed matrix obtained as in* (7) *with $\alpha = 0$ and $\widetilde{\mu}_1$ the largest eigenvalue of $\widetilde{Y}\widetilde{Y}^T$. Then, almost surely,*

- *If $\lambda > \frac{\sqrt{d_0}}{F_g}$, then $\widetilde{\mu}_1 \to (1+\lambda F_g)(1+\frac{d_0}{\lambda F_g})$.*
- *If $\lambda < \frac{\sqrt{d_0}}{F_g}$, then $\widetilde{\mu}_1 \to d_+ = (1+\sqrt{d_0})^2$.*

From Theorem 5, if $\lambda > \frac{\sqrt{d_0}}{F_g}$, we immediately see that the signal in the additive model can be reliably detected by the transformed PCA. Thus, the detection threshold in the PCA is lowered when the noise is non-Gaussian. We also remark that h_0 is the optimal entrywise transformation (up to constant factor) as in the Wigner case; see Section B.4 of Supplementary Material for more detail.

For the proof, we first adapt the strategy in (Perry et al., 2018) to justify that the transformed matrix is approximately equal to (6), which is another rectangular matrix. We then prove a BBP-type transition for the additive model, following the method of (Benaych-Georges & Nadakuditi, 2012). Since our assumptions on the spike and the noise are weaker, we provide the detail of the proof of Theorem 5 in Section B.2 of Supplementary Material.

For the multiplicative model, we have the following.

Theorem 6. *Let Y be a spiked rectangular matrix in Definition 3 that satisfy Assumption 4. Let $\widetilde{Y} \equiv \widetilde{Y}^{(\alpha_g)}$ be the transformed matrix obtained as in* (7) *with*

$$\alpha_g := \frac{-\gamma F_g + \sqrt{4F_g + 4\gamma F_g + \gamma^2 F_g^2}}{2(1+\gamma)}$$

and $\widetilde{\mu}_1$ the largest eigenvalue of $\widetilde{Y}\widetilde{Y}^T$. Then, almost surely,

- *If $\lambda_g > \sqrt{d_0}$, then $\widetilde{\mu}_1 \to (1+\lambda_g)(1+\frac{d_0}{\lambda_g})$.*
- *If $\lambda_g < \sqrt{d_0}$, then $\widetilde{\mu}_1 \to d_+ = (1+\sqrt{d_0})^2$.*

where

$$\lambda_g := \gamma + \frac{\gamma^2 F_g}{2} + \frac{\gamma\sqrt{4F_g + 4\gamma F_g + \gamma^2 F_g^2}}{2}.$$

Note that

$$\lambda_g \ge \gamma + \frac{\gamma^2 F_g}{2} + \frac{\gamma\sqrt{4 + 4\gamma F_g + \gamma^2 F_g^2}}{2} = 2\gamma + \gamma^2 F_g \\ \ge 2\gamma + \gamma^2 = \lambda,$$

and the inequality is strict if $F_g > 1$, i.e., g is not Gaussian. From Theorem 6, if $\lambda_g > \sqrt{d_0}$, we immediately see that the signal can be reliably detected by the transformed PCA. Thus, the detection threshold in the PCA is lowered when the noise is non-Gaussian. We also remark that h_{α_g} is the optimal entrywise transformation (up to constant factor); see Section B.4 of Supplementary Material for more detail.

We outline the proof of Theorem 6. We begin by justifying that the transformed matrix $\widetilde{Y}$ is approximately of the form $(Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X)$. Then, the largest eigenvalue of $\widetilde{Y}\widetilde{Y}^T$ can be approximated by the largest eigenvalue of $(Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X)^T(Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X)$ for which we consider an identity

$$\begin{aligned}&(Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X)^T(Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X) - zI\\ &= (Q^T Q - zI)(I + L(z)),\end{aligned}$$

where

$$\begin{aligned}&L(z) = \mathcal{G}(z)(\widehat{\gamma}X^T\boldsymbol{u}\boldsymbol{u}^T Q + \widehat{\gamma}Q^T\boldsymbol{u}\boldsymbol{u}^T X + \widehat{\gamma}^2 X^T\boldsymbol{u}\boldsymbol{u}^T X),\\ &\mathcal{G}(z) = (Q^T Q - zI)^{-1}.\end{aligned}$$

If z is an eigenvalue of $(Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X)^T(Q + \widehat{\gamma}\boldsymbol{u}\boldsymbol{u}^T X)$ but not of $Q^T Q$, the determinant of $(I + L(z))$ must be 0 and hence -1 is an eigenvalue of $L(z)$. Since the rank of $L(z)$ is at most 2, we can find that the eigenvector of $L(z)$ is a linear combination of two vectors $\mathcal{G}(z)Q^T\boldsymbol{u}$ and $\mathcal{G}(z)X^T\boldsymbol{u}$, i.e., for some a, b,

$$\begin{aligned}&L(z)(a\mathcal{G}(z)Q^T\boldsymbol{u} + b\mathcal{G}(z)X^T\boldsymbol{u})\\ &= -(a\mathcal{G}(z)Q^T\boldsymbol{u} + b\mathcal{G}(z)X^T\boldsymbol{u}).\end{aligned} \tag{8}$$

From the definition of $L(z)$,

$$\begin{aligned}L(z)\cdot\mathcal{G}(z)X^T\boldsymbol{u} = &\ \widehat{\gamma}\langle\boldsymbol{u}, Q\mathcal{G}(z)X^T\boldsymbol{u}\rangle\cdot\mathcal{G}(z)X^T\boldsymbol{u}\\ &+ \widehat{\gamma}\langle\boldsymbol{u}, X\mathcal{G}(z)X^T\boldsymbol{u}\rangle\cdot\mathcal{G}(z)Q^T\boldsymbol{u}\\ &+ \widehat{\gamma}^2\langle\boldsymbol{u}, X\mathcal{G}(z)X^T\boldsymbol{u}\rangle\cdot\mathcal{G}(z)X^T\boldsymbol{u},\end{aligned}$$

and a similar equation holds for $L(z)\cdot\mathcal{G}(z)Q^T\boldsymbol{u}$. It suggests that if $\langle\boldsymbol{u}, Q\mathcal{G}(z)X^T\boldsymbol{u}\rangle$ and $\langle\boldsymbol{u}, X\mathcal{G}(z)X^T\boldsymbol{u}\rangle$ are concentrated around deterministic functions of z, then the left side of (8) can be well-approximated by a (deterministic) linear combination of $\mathcal{G}(z)Q^T\boldsymbol{u}$ and $\mathcal{G}(z)X^T\boldsymbol{u}$. We can then find the location of the largest eigenvalue in terms of a deterministic function of z and conclude the proof by optimizing the function q.

The concentration of random quantities $\langle\boldsymbol{u}, Q\mathcal{G}(z)X^T\boldsymbol{u}\rangle$ and $\langle\boldsymbol{u}, X\mathcal{G}(z)X^T\boldsymbol{u}\rangle$ is the biggest technical challenge in the proof, mainly due to the dependence between the matrices Q and X. We prove it by applying the technique of linearization in conjunction with resolvent identities and also several recent results from random matrix theory, most notably the local Marchenko–Pastur law.

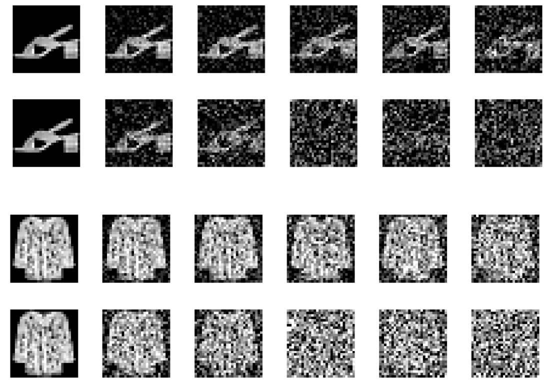

Figure 1. We compare the reconstruction performance of the proposed PCA (top lines) and the standard PCA (bottom lines) for two FashionMNIST images, with the number of measurements $N = [3136, 1568, 784, 588, 392]$ where $M = 784$. The left most column displays the original images for comparison.

The detailed proof of Theorem 6 can be found in Section B.3 of Supplementary Material.

Remark 7. Unlike the additive model, we cannot determine α_g without prior knowledge on the SNR. Nevertheless, we can apply the transformation $h_{\sqrt{F_g}}$, which effectively increases the SNR; see Section B.4 of Supplementary Material for more detail.

3.2. Applying the improved PCA to real data

To illustrate the improvement of PCA in Section 3.1, we perform the following numerical experiment: We choose a vector $\boldsymbol{z} \in \mathbb{R}^{784}$ from the standard Fashion-MNIST dataset. We then let the spike $\boldsymbol{u}$ be a normalized vector of $\boldsymbol{z}$. The j-th column of the data matrix Y is a noisy sample of the spike given by

$$Y_j = v_j\boldsymbol{u} + X_j,$$

where v_j follows Rademacher distribution and each entry of X_j is independently drawn from a centered bimodal distribution with unit variance, which is a convolution of Gaussian and Rademacher random variables, and normalized by $1/\sqrt{N}$. Our goal is to reconstruct the spike $\boldsymbol{u}$ from Y with N columns. In Fig. 1, we compare the reconstruction by the improved PCA with standard PCA over Y. With the optimal entrywise transformation, the proposed PCA outperforms the standard PCA.

While we have analyzed the improved PCA with prior information on the noise, it is possible to estimate the noise even when the noise distribution is not known. As an attempt, we tried kernel density estimation (KDE) with the Gaussian

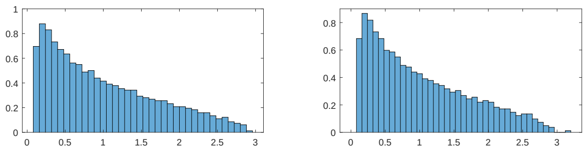

Figure 2. The spectrum of the sample covariance matrices, before (left) and after (right) the entrywise transformation $\widehat{h}$. An outlier eigenvalue can be seen only after the entrywise transformation

kernel, which approximates the density of the noise $g(x)$ by

$$\widehat{g}(x) := \frac{1}{MN\delta} \sum_{i,j} \phi((x - \sqrt{N} Y_{ij})/\delta),$$

where ϕ is the density function of the standard normal random variable and δ is the bandwidth, which we chose to be $(MN)^{-1/5}$.

For a numerical experiment, we consider the data matrix $Y = \sqrt{\lambda} \boldsymbol{u}\boldsymbol{v}^T + X$, where $\sqrt{M} u_i$ and $\sqrt{M} v_j$ follow Rademacher distribution for $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, N$. The noise is independently drawn from the same centered bimodal distribution as in the experiment above but with the variance N^{-1}. The size of the data matrix is set to be $M = 1024$, $N = 2048$, and hence the ratio $d_0 = M/N = 1/2$. We set the SNR $\lambda \approx 0.4945$. With the approximation $\widehat{g}$, we use the entrywise transformation $\widehat{h} := -\widehat{g}'/\widehat{g}$.

In Fig. 2, we compare the spectrum of the sample covariance matrices, YY^T (left) and $\widetilde{Y}\widetilde{Y}^T$ (right), where for the latter we rescale the eigenvalues so that the bulk of its spectrum matches that of the former. An isolated eigenvalue can be seen only in the spectrum in the bottom, which is the case after the entrywise transformation.

For more simulation results about the improved PCA, see Section A of Supplementary Material.

4. Main Result II - Weak Detection

In this section, we state our second main results on the hypothesis test and provide the results from numerical experiments.

4.1. Hypothesis testing and central limit theorem

Suppose that the SNR ω for the alternative hypothesis $\boldsymbol{H}_1$ is known and our goal is to detect the presence of the signal. We propose a test based on the LSS of the data matrix in (4). The key observation is that the variances of the limiting Gaussian distributions of the LSS are equal while the means are not. If we denote by $V_Y(f)$ the common variance, and $m_Y(f)|_{\boldsymbol{H}_0}$ and $m_Y(f)|_{\boldsymbol{H}_1}$ the means, respectively, our goal is to find a function that maximizes the relative difference between the limiting distributions of the LSS under $\boldsymbol{H}_0$ and under $\boldsymbol{H}_1$, i.e.,

$$\left| \frac{m_Y(f)|_{\boldsymbol{H}_1} - m_Y(f)|_{\boldsymbol{H}_0}}{\sqrt{V_Y(f)}} \right|.$$

As we will see in Theorem 11, the optimal function f is of the form $C_1 \phi_\omega + C_2$ for some constants C_1 and C_2, where

$$\begin{aligned} \phi_\omega(x) = \frac{\omega}{d_0} \left(\frac{2}{w_4 - 1} - 1 \right) x \\ - \log \left(\left(1 + \frac{d_0}{\omega} \right) (1 + \omega) - x \right). \end{aligned} \tag{9}$$

The test statistic we use is thus defined as

$$\begin{aligned} L_\omega &= \sum_{i=1}^{M} \phi_\omega(\mu_i) - M \int_{d_-}^{d_+} \phi_\omega(x) \, \mathrm{d}\mu_{MP}(x) \\ &= -\log\det \left(\left(1 + \frac{d_0}{\omega} \right) (1 + \omega) I - YY^T \right) \\ &\quad + \frac{\omega}{d_0} \left(\frac{2}{w_4 - 1} - 1 \right) (\operatorname{Tr} YY^T - M) \\ &\quad + M \left[\frac{\omega}{d_0} - \log \left(\frac{\omega}{d_0} \right) - \frac{1 - d_0}{d_0} \log(1 + \omega) \right]. \end{aligned} \tag{10}$$

Our main result in this section is the following CLT for L_ω.

Theorem 8. *Let Y be a spiked rectangular matrix in Definition 3 or 2 with $w \in (0, \sqrt{d_0})$ and $w_4 > 1$. Then, for any spikes with $\|\boldsymbol{u}\| = 1$ and $\|\boldsymbol{v}\| = 1$,*

$$L_\omega \Rightarrow \mathcal{N}(m(\lambda), V_0). \tag{11}$$

The mean of the limiting Gaussian distribution is given by

$$\begin{aligned} m(\lambda) = -\frac{1}{2} \log \left(1 - \frac{\omega^2}{d_0} \right) - \frac{\omega^2}{2d_0}(w_4 - 3) \\ - \log \left(1 - \frac{\lambda^2}{d_0} \right) + \frac{\lambda^2}{d_0} \left(\frac{2}{w_4 - 1} - 1 \right) \end{aligned} \tag{12}$$

and the variance

$$V_0 = -2 \log \left(1 - \frac{\omega^2}{d_0} \right) + \frac{2\omega^2}{d_0} \left(\frac{2}{w_4 - 1} - 1 \right) \tag{13}$$

Theorem 8 directly follows from the general CLT result in Theorem 11. See also Section C.3 of Supplementary Material for more detail on the mean and the variance.

We propose a test in Algorithm 1 based on Theorem 8. In this test, we compute the test statistic L_ω and compare it

Algorithm 1 Hypothesis test

Input: data Y_{ij}, parameters w_4, ω
$L_\omega \leftarrow$ test statistic in (10)
$m_\omega \leftarrow$ critical value in (14)
if $L_\omega \le m_\omega$ **then**
  Accept $\boldsymbol{H}_0$
else
  Reject $\boldsymbol{H}_0$
end if

with the average of $m(0)$ and $m(\omega)$, i.e.,

$$m_\omega := \frac{m(0)+m(\omega)}{2} = -\log\left(1-\frac{\omega^2}{d_0}\right) + \frac{\omega^2}{2d_0}\left(\frac{2}{w_4-1} - w_4 + 2\right). \tag{14}$$

As a simple corollary to Theorem 8, we have the following formula for the limiting error of the proposed test.

Corollary 9. *The error of the test in algorithm 1,*

$$\mathrm{err}(\omega) = \mathbb{P}(L_\omega > m_\omega | \boldsymbol{H}_0) + \mathbb{P}(L_\omega \le m_\omega | \boldsymbol{H}_1) \to \mathrm{erfc}\left(\frac{\sqrt{V_0}}{4\sqrt{2}}\right), \tag{15}$$

where V_0 is the variance in (13) *and* $\mathrm{erfc}(\cdot)$ *is the complementary error function.*

For the proof of Corollary 9, see Corollary 5 of (El Alaoui et al., 2020) or Theorem 2 of (Chung & Lee, 2019).

Remark 10. If the noise X is Gaussian, $w_4 = 3$ and the limiting error in Corollary 9 is

$$\mathrm{erfc}\left(\frac{\sqrt{V_0}}{4\sqrt{2}}\right) = \mathrm{erfc}\left(\frac{1}{4}\sqrt{-\log\left(1-\frac{\omega^2}{d_0}\right)}\right),$$

and it coincides with the error of the LR test in (3). It shows that our test is optimal with the Gaussian noise.

Even if the exact parameter w_4 is not known a priori, it can be easily estimated from the data matrix Y by computing $\frac{1}{MN}\sum Y_{ij}^4$. The accuracy of such an estimate can be easily checked from the Chernoff bound.

Lastly, we state a general CLT for the LSS and the optimality of the function ϕ_ω as the test statistic.

Theorem 11. *Assume the conditions in Theorem 8. Denote by $\mu_1 \ge \mu_2 \ge \cdots \ge \mu_M$ the eigenvalues of YY^T. For any function f analytic on an open set containing an interval $[d_-, d_+]$,*

$$\left(\sum_{i=1}^M f(\mu_i) - M\int_{d_-}^{d_+} \frac{\sqrt{(x-d_-)(d_+-x)}}{2\pi d_0 x} f(x)\,\mathrm{d}x\right) \Rightarrow \mathcal{N}(m_Y(f), V_Y(f)). \tag{16}$$

The mean and the variance of the limiting Gaussian distribution are given by

$$m_Y(f) = \frac{\widetilde{f}(2)+\widetilde{f}(-2)}{4} - \frac{\tau_0(\widetilde{f})}{2} - (w_4-3)\tau_2(\widetilde{f}) + \sum_{\ell=1}^\infty \left(\frac{\omega}{\sqrt{d_0}}\right)^\ell \tau_\ell(\widetilde{f})$$

and

$$V_Y(f) = 2\sum_{\ell=1}^\infty \ell\tau_\ell(\widetilde{f})^2 + (w_4-3)\tau_1(\widetilde{f})^2,$$

where we let $\widetilde{f}(x) = f(\sqrt{d_0}x + 1 + d_0)$,

$$\tau_\ell(f) = \frac{1}{\pi}\int_{-2}^2 T_\ell\left(\frac{x}{2}\right)\frac{f(x)}{\sqrt{4-x^2}}\mathrm{d}x,$$

and T_ℓ is the ℓ-th Chebyshev polynomial of the first kind.

Furthermore, for $m(\omega)$, $m(0)$, and V_0 defined in Theorem 8,

$$\left|\frac{m_Y(f)-m_X(f)}{\sqrt{V_Y(f)}}\right| \le \left|\frac{m(\omega)-m(0)}{\sqrt{V_0}}\right|$$

The equality holds if and only if $f(x) = C_1\phi_\omega(x) + C_2$ for some constants C_1 and C_2 with the function ϕ_ω defined in (9).

We remark that the analyticity of the function f in Theorem 11 is assumed only because it is sufficient in our purpose and this assumption can be weakened by the density argument, which is typically used in the proof of CLT results in random matrix theory.

We briefly sketch the proof of Theorem 11 based on the interpolation technique, developed in (Chung & Lee, 2019; Jung et al., 2020). In this method, the right side of (16) is written as the following contour integral of the trace of the resolvent: For a function f analytic on an open set containing an interval $[d_-, d_+]$,

$$\sum_{i=1}^M f(\mu_i) = \sum_{i=1}^M \frac{1}{2\pi\mathrm{i}}\oint_\Gamma \frac{f(z)}{z-\mu_i}\mathrm{d}z = -\frac{1}{2\pi\mathrm{i}}\oint_\Gamma f(z)\,\mathrm{Tr}(YY^T - zI)^{-1}\mathrm{d}z \tag{17}$$

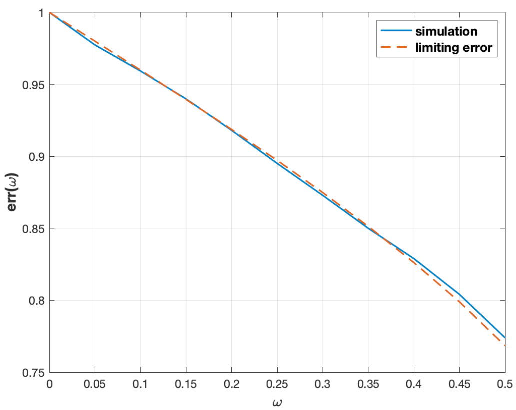

Figure 3. The error from the simulation (solid) and the theoretical limiting error in (15) (dashed), respectively, for the Gaussian noise.

for any contour Γ containing $\mu_1, \mu_2, \ldots, \mu_N$. For the null model, i.e., if $\lambda = 0$, the CLT was proved in (Bai & Silverstein, 2004; Lytova & Pastur, 2009) with precise formulas for the mean and the variance.

To prove the CLT for a non-null model, i.e., a spiked rectangular matrix with $\lambda \neq 0$, we introduce an interpolation between the null model and the non-null model, and track the change of the LSS by finding the change of $\mathrm{Tr}(YY^T - zI)^{-1}$. The change is decomposed into the deterministic part and the random part, where the latter converges to 0 with overwhelming probability for both the additive model and the multiplicative model. We can then conclude that the CLT for the LSS holds also for the non-null model, with the variance unchanged. The change of the mean can be computed by considering the deterministic change of the resolvent.

The detail of the proof of Theorem 11 can be found in Section C of Supplementary Material.

4.2. Numerical experiments for the LSS test

We consider the case where the noise matrix X is Gaussian and the signal $\boldsymbol{u} = (u_1, u_2, \ldots, u_M)^T$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_N)^T$, where $\sqrt{M}u_i$'s and $\sqrt{N}v_j$'s are i.i.d. Rademacher random variables for $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$. Let the data matrix $Y = \sqrt{\lambda}\boldsymbol{u}\boldsymbol{v}^T + X$. The parameters are $w_2 = 2$ and $w_4 = 3$.

In Figure 3, we plot empirical average (after 10,000 Monte Carlo simulations) of the error of the proposed test in Algorithm 1 and the theoretical (limiting) error in (15), varying the SNR ω from 0 to 0.5, with $M = 256$ and $N = 512$. It can be checked that the error of the proposed test closely matches the theoretical error.

5. Conclusion and Future Works

In this paper, we considered the detection problem of spiked rectangular model. For both the multiplicative model and the additive model, we showed that PCA can be improved for non-Gaussian noise by transforming the data entrywise. We proved the effective SNR and the optimal entrywise transforms for both models. We also proposed a universal test that does not require any prior information on the spike. The test and its error do not depend on the noise except its (normalized) fourth moment. The error of the proposed test is optimal when the noise is Gaussian.

A natural future research direction is to apply the entrywise transformation for the weak detection. As in the spiked Wigner model, we believe that the error of the proposed test can be lowered with the entrywise transformation and it can be proved by establishing the central limit theorems for the transformed matrices.

Acknowledgements

The authors would like to thank anonymous reviewers for their suggestions and feedback. The work of J. H. Jung and J. O. Lee was partially supported by National Research Foundation of Korea under grant number NRF-2019R1A5A1028324. The work of H. W. Chung was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2017R1E1A1A01076340 and No. 2021R1C1C1008539).

References

Abbe, E. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Bai, Z. D. and Silverstein, J. W. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605, 2004.

Baik, J. and Lee, J. O. Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model. *J. Stat. Phys.*, 165(2):185–224, 2016.

Baik, J. and Lee, J. O. Free energy of bipartite spherical Sherrington–Kirkpatrick model. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(4):2897–2934, 2020.

Baik, J., Ben Arous, G., and Péché, S. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

Benaych-Georges, F. and Nadakuditi, R. R. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, 227(1):494–521, 2011.

Benaych-Georges, F. and Nadakuditi, R. R. The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.*, 111: 120–135, 2012.

Bloemendal, A., Knowles, A., Yau, H.-T., and Yin, J. On the principal components of sample covariance matrices. *Probab. Theory Related Fields*, 164(1-2):459–552, 2016.

Chung, H. W. and Lee, J. O. Weak detection of signal in the spiked wigner model. In *International Conference on Machine Learning*, pp. 1233–1241, 2019.

Dobriban, E. Sharp detection in PCA under correlations: All eigenvalues matter. *Ann. Statist.*, 45(4):1810 – 1833, 2017.

El Alaoui, A. and Jordan, M. I. Detection limits in the high-dimensional spiked rectangular model. In *Conference On Learning Theory*, pp. 410–438, 2018.

El Alaoui, A., Krzakala, F., and Jordan, M. I. Fundamental limits of detection in the spiked Wigner model. *Ann. Statist.*, 48(2):863–885, 2020.

Johnstone, I. M. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2): 295–327, 2001.

Johnstone, I. M. High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I*, pp. 307–333. Eur. Math. Soc., Zürich, 2007.

Jung, J. H., Chung, H. W., and Lee, J. O. Weak detection in the spiked wigner model with general rank. *arXiv:2001.05676*, 2020.

Lesieur, T., Krzakala, F., and Zdeborová, L. Mmse of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 680–687. IEEE, 2015.

Lytova, A. and Pastur, L. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37(5):1778–1840, 2009.

Montanari, A., Reichman, D., and Zeitouni, O. On the limitation of spectral methods: from the Gaussian hidden clique problem to rank one perturbations of Gaussian tensors. *IEEE Trans. Inform. Theory*, 63(3):1572–1579, 2017.

Montanari, A., Ruan, F., and Yan, J. Adapting to unknown noise distribution in matrix denoising. *arXiv:1810.02954*, 2018.

Onatski, A., Moreira, M. J., and Hallin, M. Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.*, 41(3):1204–1231, 2013.

Onatski, A., Moreira, M. J., and Hallin, M. Signal detection in high dimension: the multispiked case. *Ann. Statist.*, 42 (1):225–254, 2014.

Perry, A., Wein, A. S., Bandeira, A. S., and Moitra, A. Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.*, 46(5):2416–2451, 2018.