Weak Detection of Signal in the Spiked Wigner Model

Hye Won Chung [1] **Ji Oon Lee** [2]

Abstract

We consider the problem of detecting the presence of the signal in a rank-one signal-plus-noise data matrix. In case the signal-to-noise ratio is under the threshold below which a reliable detection is impossible, we propose a hypothesis test based on the linear spectral statistics of the data matrix. When the noise is Gaussian, the error of the proposed test is optimal as it matches the error of the likelihood ratio test that minimizes the sum of the Type-I and Type-II errors. The test is data-driven and does not depend on the distribution of the signal or the noise. If the density of the noise is known, it can be further improved by an entrywise transformation to lower the error of the test.

1. Introduction

One of the fundamental questions in statistics is to detect signals from given data. When the data is given as a matrix, it is common to analyze the data by the largest eigenvalue and the corresponding eigenvector, which is known as principal component analysis (PCA). For a null model where the signal is not present, the data is pure noise and the behavior of the largest eigenvalue is now well understood by random matrix theory (Tracy & Widom, 1994; 1996; Johnstone, 2001; Tao & Vu, 2010; Erdős et al., 2012). If the data matrix is of 'signal-plus-noise' type and the signal is in the form of a vector, the model is often referred to as a 'spiked random matrix.'

When the signal is an N-dimensional vector and the data is an $N \times N$ real symmetric matrix, one of the most natural signal-plus-noise models is of the form

$$M = \sqrt{\lambda}\boldsymbol{x}\boldsymbol{x}^T + H, \tag{1}$$

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

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

where the signal $\boldsymbol{x} \in \mathbb{R}^N$ and H is an $N \times N$ Wigner matrix. (See Definitions 1 and 2.) The spiked Wigner model is widely used as a low-rank model for which PCA can be applied to detect or recover signal from noisy high dimensional data. It can be used in the signal detection/recovery problems such as community detection (Abbe, 2017) and submatrix localization (Butucea et al., 2013). In the community detection, the spike $\boldsymbol{x} \in \{1, -1\}^N$ indicates communities each node belongs to and the data matrix M models noisy pairwise interactions with different means depending on whether the corresponding nodes are in the same community or not. In the submatrix localization, the task is to detect within a large Gaussian matrix small blocks with atypical mean.

The Wigner matrix H represents the noise, and we assume that H_{ij} are independent random variables with mean 0 and variance N^{-1}. With the assumption, the spectral norm of H, $\|H\| \to 2$ as $N \to \infty$ almost surely. Thus, when the signal is normalized so that $\|\boldsymbol{x}\|_2 = 1$, the strengths of the signal $\|\boldsymbol{x}\boldsymbol{x}^T\|$ and the noise $\|H\|$ are comparable. If the parameter λ, which corresponds to the signal-to-noise ratio (SNR), is very large ($\lambda \gg 1$) or small ($\lambda \ll 1$), a perturbation argument can be applied for PCA; if $\lambda \gg 1$, the difference between the largest eigenvalues of M and $\sqrt{\lambda}\boldsymbol{x}\boldsymbol{x}^T$ is negligible, and if $\lambda \ll 1$, the largest eigenvalue of M cannot be distinguished from that of H.

The case $\lambda \sim 1$ has been intensively studied in random matrix theory. The first result in this direction was obtained by Baik, Ben Arous, and Péché (Baik et al., 2005) for complex Wishart matrices, which is of the form X^*X where X is a (rectangular) matrix with independent Gaussian entries, and later extended to more general sample covariance matrices (Paul, 2007; Nadler, 2008; Johnstone & Lu, 2009). Similar results were proved for Wigner matrices (Péché, 2006; Féral & Péché, 2007; Capitaine et al., 2009; Benaych-Georges & Nadakuditi, 2011). In these results, the largest eigenvalue exhibits phase transition; when $\lambda > 1$, the largest eigenvalue separates from the other eigenvalues of M and converges to $\sqrt{\lambda} + \frac{1}{\sqrt{\lambda}}$, which is strictly larger than 2, whereas for $\lambda < 1$, the behavior of the largest eigenvalue coincides with that of the pure noise model. In the former case, the eigenvector corresponding to the largest eigenvalue has nontrivial correlation with the signal $\boldsymbol{x}$, which means that the signal can be detected and recovered by PCA. We refer to the work of Benaych-Georges and Nadakuditi (Benaych-Georges & Nadakuditi, 2011) for more detail on the behavior of the largest eigenvalues and corresponding eigenvectors.

When $\lambda < 1$, contrary to the case $\lambda > 1$, the spectral norm of M converges to 2 and the behavior of the largest eigenvalue cannot be distinguished from that of the null model H. It is then natural to ask whether the presence of the signal is detectable, and if so, which tests allow us to detect the signal in the regime $\lambda < 1$.

The question about the detectability was considered by Montanari, Reichman, and Zeitouni in (Montanari et al., 2017), where it was proved that no tests based on the eigenvalues can reliably detect the signal if the noise H is a random matrix from the Gaussian Orthogonal Ensemble (GOE). For a non-Gaussian Wigner matrix H, Perry, Wein, Bandeira, and Moitra (Perry et al., 2018) assumed that the signal $\boldsymbol{x}$ is drawn from a distribution $\mathcal{X}$, which they called the spike prior, and found the critical value for $\lambda \leq 1$ in terms of $\mathcal{X}$ and H below which no tests based on the eigenvalues can reliably detect signal. Further, they also established an entrywise transformation of the data matrix by which the signal can be detected via the largest eigenvalue even if $\lambda < 1$ as long as λ is larger than the critical value.

For the subcritical case, El Alaoui, Krzakala, and Jordan (El Alaoui et al., 2018) studied the weak detection, i.e., a test with accuracy better than a random guess. More precisely, they considered the hypothesis testing problem between the null hypothesis that $\lambda = 0$ and the alternative hypothesis that M is generated with a fixed $\lambda > 0$. Assuming that the entries of $\sqrt{N}\boldsymbol{x}$ are i.i.d. random variables with bounded support and the noise is Gaussian, it was proved that the error from the likelihood ratio (LR) test, which is the optimal test in minimizing the error, converges to

$$\operatorname{erfc}\left(\frac{1}{4}\sqrt{-\log(1-\lambda)-\lambda}\right) \tag{2}$$

if the variance of diagonal entries H_{ii} tends to infinity. We remark that Onatski, Moreira, and Hallin (Onatski et al., 2013) considered the weak detection for real Wishart matrices and obtained a Gaussian limit of the log LR.

While the likelihood ratio test is optimal due to the Neyman–Pearson lemma, since LR tests require substantial knowledge of the distribution of x, called prior, it is desirable to design a test that does not require a priori knowledge on the signal. For community detection problem in the stochastic block model, Banerjee and Ma (Banerjee & Ma, 2017) proposed a test based on the linear spectral statistics (LSS). More precisely, denoting by $\mu_1, \dots, \mu_N$ the eigenvalues of the data matrix, they considered the LSS

$$L_N(f) = \sum_{i=1}^{N} f(\mu_i) \tag{3}$$

with $f(x) = x^k$ for positive integers k and achieved asymptotically optimal error by a linear combination of the LSS.

The results in (El Alaoui et al., 2018; Banerjee & Ma, 2017) shed lights on the weak detection problem. However, the analysis in these results seems to be restricted to the specific distributions of the noise - Gaussian distribution in (El Alaoui et al., 2018) and Bernoulli distribution in (Banerjee & Ma, 2017). Moreover, the signal considered in the previous works is delocalized, i.e., $\|\boldsymbol{x}\|_\infty = O(1/\sqrt{N})$, which may lose its validity if the signal is sparse or the uniform prior is assumed.

In this paper, we construct an optimal and universal test that detects the absence or presence of signal in (1) based on LSS for any $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_2 = 1$ and for any Wigner matrix H. We briefly summarize our main contributions as follows:

- Universality 1: For any deterministic or random $\boldsymbol{x}$, the proposed test and its error do not change, and thus we do not need any prior information on $\boldsymbol{x}$. Note that the LR test requires the prior information on $\boldsymbol{x}$.
- Universality 2: The proposed test and its error depend on the distribution of the noise H only through the variance of the diagonal entries and the fourth moment of the off-diagonal entries. The entries H_{ij} do not need to be identically distributed but just independent.
- Optimality 1: The proposed test is with the lowest error among all tests based on LSS.
- Optimality 2: When the noise is Gaussian, the error of the proposed test with low computational complexity converges to the optimal limit (2) obtained in (El Alaoui et al., 2018).
- Data-driven test: The various quantities in the proposed test can be estimated from the observed data.
- Entrywise transformation: If the density function of the noise matrix is known, which is non-Gaussian, the test can be further improved by an entrywise transformation that effectively increases the SNR.

The main technical component of the present paper is the central limit theorem (CLT) for the LSS of arbitrary analytic functions for the random matrix in (1). The fluctuation of the LSS is not only of fundamental importance per se in random matrix theory, but also applicable to various applications such as the fluctuations of the free energy of the spherical spin glass (Baik & Lee, 2016; 2017). The LR in the weak detection problem with Gaussian noise is directly related to the free energy of spin glass as in (El Alaoui et al., 2018). To our best knowledge, however, the CLT for spiked Wigner matrices was proved only for the case where the signal $\boldsymbol{x} = \mathbf{1} := \frac{1}{\sqrt{N}}(1, 1, \dots, 1)^T$ (Baik & Lee, 2017).

If the density function of the noise matrix is known, we can adapt the entrywise transformation in (Perry et al., 2018) to further improve the proposed test. The transformation effectively changes the SNR from λ to λF^H, where F^H is the Fisher information of the density function of the (normalized) off-diagonal entries in H. Since $F^H \geq 1$, and the strict inequality holds if the noise is non-Gaussian, the error from our test decreases in general after the transformation.

The rest of the paper is organized as follows. In Section 2, we define the model and introduce previous results. In Section 3, we state the main result and describe the algorithm for the proposed test. In Section 4, we apply the entrywise transformation and state the results for the improved test. General results on the CLT for the LSS are collected in Section 5. We conclude the paper in Section 6 with the summary of our results and future research directions. Proof of the theorems can be found in Supplementary Material.

2. Preliminaries

We first define the matrix in (1) more precisely. The Wigner matrix is defined as follows:

Definition 1 (Wigner matrix)**.** *We say an $N \times N$ random matrix $H = (H_{ij})$ is a (real) Wigner matrix if H is a symmetric matrix and H_{ij} $(1 \leq i \leq j \leq N)$ are independent real random variables satisfying the following conditions:*

- *All moments of H_{ij} are finite and $\mathbb{E}[H_{ij}] = 0$.*
- *For all $i < j$, $N\mathbb{E}[H_{ij}^2] = 1$, $N^{\frac{3}{2}}\mathbb{E}[H_{ij}^3] = w_3$, $N^2\mathbb{E}[H_{ij}^4] = w_4$ for some constants $w_3, w_4 \in \mathbb{R}$.*
- *For all i, $N\mathbb{E}[H_{ii}^2] = w_2$ for a constant $w_2 \geq 0$.*

The signal-plus-noise model we consider is a (rank-one) spiked Wigner matrix, which is defined as follows:

Definition 2 (Spiked Wigner matrix)**.** *We say an $N \times N$ random matrix $M = \sqrt{\lambda}\boldsymbol{x}\boldsymbol{x}^T + H$ is a spiked Wigner matrix with a spike $\boldsymbol{x}$ and SNR λ if $\boldsymbol{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ with $\|\boldsymbol{x}\|_2 = 1$ and H is a Wigner matrix.*

Denote by $\mathbb{P}_\lambda$ the joint probability of the observation, a spiked Wigner matrix, with $\lambda > 0$ and $\mathbb{P}_0$ with $\lambda = 0$. If H is a GOE matrix, where H_{ij} are Gaussian with $N\mathbb{E}[H_{ii}^2] = 2$, and $\boldsymbol{x}$ is drawn from the spike prior $\mathcal{X}$, the likelihood ratio is given by

$$\frac{\mathrm{d}\mathbb{P}_\lambda}{\mathrm{d}\mathbb{P}_0} = \int \exp\Big(\frac{N}{2}\sum_{i,j=1}^{N}\big(\sqrt{\lambda}M_{ij}x_ix_j - \frac{\lambda}{2}x_i^2x_j^2\big)\Big)\mathrm{d}\mathcal{X}(\boldsymbol{x}).$$

For the spherical prior, i.e., $\mathcal{X}$ is the uniform distribution on the unit sphere, with the spike $\boldsymbol{x} = \boldsymbol{1}$, it was proved that

$$\log\frac{\mathrm{d}\mathbb{P}_\lambda}{\mathrm{d}\mathbb{P}_0} \Rightarrow \mathcal{N}\left(\pm\frac{1}{4}\log\left(\frac{1}{1-\lambda}\right), \frac{1}{4}\log\left(\frac{1}{1-\lambda}\right)\right),$$

where the plus sign holds under the alternative $M \sim \mathbb{P}_\lambda$ and the minus sign holds under the null $M \sim \mathbb{P}_0$. (See Section 3.1 of (Baik & Lee, 2016) and Theorem 1.4 of (Baik & Lee, 2017) with $\beta = \sqrt{\lambda}/2$.) For the i.i.d. bounded prior, i.e., the entries of $\sqrt{N}\boldsymbol{x}$ are i.i.d. bounded random variables, the same result was proved in (El Alaoui et al., 2018).

The proof of the convergence of $\frac{\mathrm{d}\mathbb{P}_\lambda}{\mathrm{d}\mathbb{P}_0}$ in (Baik & Lee, 2016; 2017) is based on the recent development of random matrix theory, especially the study of the LSS. For a Wigner matrix H, if we let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the eigenvalues of H, then for any continuous function f defined on a neighborhood of $[-2, 2]$,

$$\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i) \to \int_{-2}^{2}\frac{\sqrt{4-x^2}}{2\pi}f(x)\,\mathrm{d}x$$

almost surely. The fluctuation of $\frac{1}{N}\sum_i f(\lambda_i)$ about its limit is a subject of intensive study in random matrix theory, and it is natural to introduce the LSS defined in (3) for the analysis. The CLT for the LSS asserts that

$$\left(L_N(f) - N\int_{-2}^{2}\frac{\sqrt{4-x^2}}{2\pi}f(x)\,\mathrm{d}x\right) \Rightarrow \mathcal{N}(m_H(f), V_H(f)), \tag{4}$$

where the right-hand side denotes a Gaussian random variable with mean $m_H(f)$ and variance $V_H(f)$. Note that the fluctuation is of order N^{-1} and much smaller than that of the conventional CLT, which is of order $N^{-\frac{1}{2}}$.

For spiked Wigner matrices, the CLT for the LSS has been proved only for the case $\boldsymbol{x} = \boldsymbol{1} := \frac{1}{\sqrt{N}}(1, 1, \dots, 1)^T$ in (Baik & Lee, 2017). Let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_N$ be the eigenvalues of a spiked Wigner matrix with a spike $\boldsymbol{x}$ and SNR λ. If $\boldsymbol{x} = \boldsymbol{1}$, then

$$\left(\sum_{i=1}^{N} f(\mu_i) - N\int_{-2}^{2}\frac{\sqrt{4-x^2}}{2\pi}f(x)\,\mathrm{d}x\right) \Rightarrow \mathcal{N}(m_M(f), V_M(f)), \tag{5}$$

A remarkable fact in (5) is that the variance $V_M(f)$ is equal to $V_H(f)$, the variance from the Wigner case, whereas the mean $m_M(f)$ is different from $m_H(f)$ unless $\lambda = 0$. (See Theorem 5 in Section 5 for the precise formulas for $m_M(f)$ and $V_M(f)$.) It turns out that the same CLT holds for any spike $\boldsymbol{x}$ as in Theorem 5, and the LSS provides us a test statistic for a hypothesis testing.

3. Main Results

Let us denote by $\boldsymbol{H}_0$ the null hypothesis and $\boldsymbol{H}_1$ the alternative hypothesis, i.e.,

$$\boldsymbol{H}_0 : \lambda = 0, \qquad \boldsymbol{H}_1 : \lambda > 0.$$

Suppose that the value of λ for $\boldsymbol{H}_1$ is known and our task is to detect whether the signal is present from a given data matrix M. If we construct a test based on the LSS for the hypothesis testing, it is obvious that we need to maximize

$$\left| \frac{m_M(f) - m_H(f)}{\sqrt{V_M(f)}} \right| . \tag{6}$$

In Theorem 6 in Section 5, we prove that the maximum of (6) is attained if and only if $f(x) = C_1 \phi_\lambda(x) + C_2$ for some constants C_1 and C_2, where

$$\phi_\lambda(x) := \log\left(\frac{1}{1-\sqrt{\lambda}x+\lambda}\right) + \sqrt{\lambda}\left(\frac{2}{w_2}-1\right)x + \lambda\left(\frac{1}{w_4-1}-\frac{1}{2}\right)x^2. \tag{7}$$

Thus, it is natural to define the test statistic L_λ by

$$\begin{aligned} L_\lambda &:= \sum_{i=1}^N \phi_\lambda(\mu_i) - N\int_{-2}^{2} \frac{\sqrt{4-z^2}}{2\pi}\phi_\lambda(z)\,\mathrm{d}z \\ &= -\log\det\left((1+\lambda)I - \sqrt{\lambda}M\right) + \frac{\lambda N}{2} \\ &\quad + \sqrt{\lambda}\left(\frac{2}{w_2}-1\right)\operatorname{Tr} M \\ &\quad + \lambda\left(\frac{1}{w_4-1}-\frac{1}{2}\right)(\operatorname{Tr} M^2 - N). \end{aligned} \tag{8}$$

Under $\boldsymbol{H}_0$, it is direct to see from Theorem 1.1 of (Bai & Yao, 2005) and Section 3.1 of (Baik & Lee, 2016) that

$$L_\lambda \Rightarrow \mathcal{N}(m_0, V_0),$$

where

$$m_0 = -\frac{1}{2}\log(1-\lambda) + \left(\frac{w_2-1}{w_4-1}-\frac{1}{2}\right)\lambda + \frac{(w_4-3)\lambda^2}{4}, \tag{9}$$

$$V_0 = -2\log(1-\lambda) + \left(\frac{4}{w_2}-2\right)\lambda + \left(\frac{2}{w_4-1}-1\right)\lambda^2. \tag{10}$$

Our first main result is the CLT for L_λ under $\boldsymbol{H}_1$.

Theorem 1. *Let M be a spiked Wigner matrix in Definition 2 with $0 < \lambda < 1$. Denote by $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_N$ the eigenvalues of M. Then, for any spike $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_2 = 1$,*

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

The mean of the limiting Gaussian distribution is given by

$$m_+ = m_0 - \log(1-\lambda) + \left(\frac{2}{w_2}-1\right)\lambda + \left(\frac{1}{w_4-1}-\frac{1}{2}\right)\lambda^2 \tag{12}$$

and the variance V_0 is as in (10).

Theorem 1 is a direct consequence of a general CLT in Theorem 5 in Section 5. (See Supplementary Material.)

In the simplest case with $w_2 = 2$ and $w_4 = 3$, e.g., when H is a GOE matrix,

$$L_\lambda \Rightarrow \mathcal{N}(-\frac{1}{2}\log(1-\lambda), -2\log(1-\lambda)) \tag{13}$$

under $\boldsymbol{H}_0$ and

$$L_\lambda \Rightarrow \mathcal{N}(-\frac{3}{2}\log(1-\lambda), -2\log(1-\lambda)) \tag{14}$$

under $\boldsymbol{H}_1$ as shown in Figure 1.

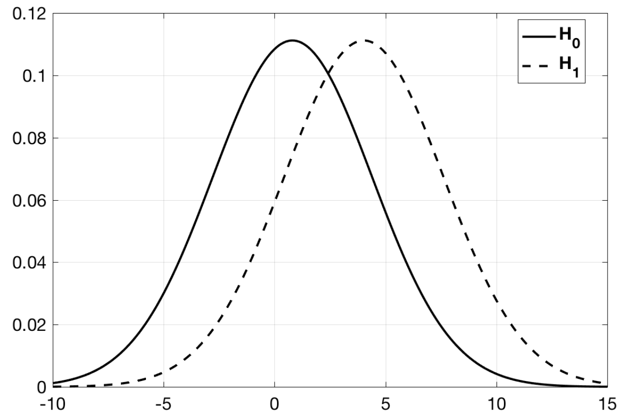

Figure 1. The limiting density of the test statistic L_λ in (13) and (14) under $\boldsymbol{H}_0$ (solid) and under $\boldsymbol{H}_1$ (dashed), respectively, with $\lambda = 0.8$ for $w_2 = 2$ and $w_4 = 3$ (e.g. GOE noise)

Based on Theorem 1, we propose a hypothesis test described in Algorithm 1. In this test, given a data matrix M, we compute L_λ and compare it with the critical value

$$\begin{aligned} m_\lambda &:= \frac{m_0 + m_+}{2} \\ &= -\log(1-\lambda) + (w_2 - 1)\left(\frac{1}{w_4-1}-\frac{1}{w_2}\right)\lambda \\ &\quad + \left(\frac{w_4}{4} - 1 + \frac{1}{2(w_4-1)}\right)\lambda^2 \end{aligned} \tag{15}$$

to accept or reject the null hypothesis test.

Theorem 2. *The error of the test in algorithm 1,*

$$\operatorname{err}(\lambda) = \mathbb{P}(L_\lambda > m_\lambda | \boldsymbol{H}_0) + \mathbb{P}(L_\lambda \leq m_\lambda | \boldsymbol{H}_1),$$

converges to $\operatorname{erfc}(E_\lambda/4)$*, where*

$$E_\lambda^2 = \log\left(\frac{1}{1-\lambda}\right) + \left(\frac{2}{w_2}-1\right)\lambda + \left(\frac{1}{w_4-1}-\frac{1}{2}\right)\lambda^2$$

and $\operatorname{erfc}(\cdot)$ *is the complementary error function defined as*

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty e^{-x^2}\,\mathrm{d}x\,.$$

Algorithm 1 Hypothesis test

Input: data M_{ij}, parameters w_2, w_4, λ
$L_\lambda \leftarrow$ test statistic in (8)
$m_\lambda \leftarrow$ critical value in (15)
if $L_\lambda \le m_\lambda$ **then**
Accept $\boldsymbol{H}_0$
else
Reject $\boldsymbol{H}_0$
end if

Proof. Due to the symmetry, $\mathbb{P}(L_\lambda > m_\lambda | \boldsymbol{H}_0)$ and $\mathbb{P}(L_\lambda \le m_\lambda | \boldsymbol{H}_1)$ converge to a common limit. Since

$$\mathbb{P}(L_\lambda > m_\lambda | \boldsymbol{H}_0) = \mathbb{P}\left(\frac{L_\lambda - m_0}{\sqrt{V_0}} \ge \frac{m_+ - m_0}{2\sqrt{V_0}} \middle| \boldsymbol{H}_0 \right)$$

and $V_0 = 2(m_+ - m_0)$, we can identify the limit as

$$\mathbb{P}\left(Z \ge \frac{\sqrt{V_0}}{4} \right)$$

for a standard Gaussian random variable Z. Thus, we can conclude that

$$\lim_{N\to\infty} \operatorname{err}(\lambda) = 2\mathbb{P}\left(Z \ge \frac{\sqrt{V_0}}{4} \right) = \operatorname{erfc}\left(\frac{E_\lambda}{4} \right). \quad (16)$$

□

Remark 1. Even when the exact values of w_2 and w_4 are not known, we can estimate the parameters from the data matrix. Such estimates are accurate enough for the algorithm as we can easily check from the Chernoff bound.

In case $w_4 = 3$ and $w_2 = \infty$, we obtain

$$\lim_{N\to\infty} \operatorname{err}(\lambda) = \operatorname{erfc}\left(\frac{1}{4}\sqrt{-\log(1-\lambda) - \lambda} \right), \quad (17)$$

which is equal to the error of the LR test, given in Corollary 5 of (El Alaoui et al., 2018). Furthermore, in case $w_4 = 3$ and $w_2 < \infty$, we get

$$\lim_{N\to\infty} \operatorname{err}(\lambda) = \operatorname{erfc}\left(\frac{1}{4}\sqrt{-\log(1-\lambda) - \lambda + \frac{2\lambda}{w_2}} \right), \quad (18)$$

which coincides with the error of the LR test, obtained in the remark after Theorem 2 of (El Alaoui et al., 2018) with $\mathbb{E}_{P_X}[X^3] = 0$. Thus, our test achieves the optimal error when λ is below a certain threhold λ_c above which reliable detection is possible as shown in (Barbier et al., 2016; Lelarge & Miolane, 2019; El Alaoui et al., 2018). Since $\lambda_c = 1$ in many cases including spherical, Rademacher, and any i.i.d. prior with a sub-Gaussian bound (Perry et al., 2018), and since universal tests such as ours cannot exploit the knowledge on the prior, we have considered a test that works for any $\lambda < 1$.

4. Test with Entrywise Transformation

In this section, we consder the case the density function of the noise matrix is known, and we further improve the proposed test by adapting the entrywise transformation in (Perry et al., 2018). Suppose that each normalized entry $\sqrt{N}H_{ij}$ is drawn from a distribution $\mathcal{P}$ with a density function g. As shown in (Perry et al., 2018), it turns out that the signal can be reliably detected by PCA if $\lambda > 1/F^H$, where F^H is the Fisher information of $\mathcal{P}$ defined by

$$F^H = \int_{-\infty}^{\infty} \frac{g'(w)^2}{g(w)} \mathrm{d}w. \quad (19)$$

Note that $F^H \ge 1$ with equality if and only if $\mathcal{P}$ is a Gaussian. The main idea of improving the detection threshold for PCA is based on the following transformation. Set

$$h(w) := -\frac{g'(w)}{g(w)}.$$

Given the data matrix M, one can consider a transformed matrix $\widetilde{M}$ obtained by

$$\widetilde{M}_{ij} = \frac{1}{\sqrt{F^H N}} h(\sqrt{N} M_{ij}).$$

The transformation effectively changes the SNR from λ to λF^H for PCA, and thus it is possible to reliably detect the signal if $\lambda F^H > 1$. For more detail, see Section 4 of (Perry et al., 2018).

If $\lambda < 1/F^H$, no tests based on PCA are reliable. Hence, we consider the weak detection of the signal with the entrywise transformation. The effective change of the SNR by the entrywise transformation suggests that the result in Theorem 1 will also change correspondingly with the entrywise transformation. For analysis, we will assume the following:

Assumption 1. *For the spike $\boldsymbol{x}$, we assume that $\|\boldsymbol{x}\|_\infty \le N^{-\phi}$ for some $\phi > \frac{3}{8}$.*

For the noise, let $\mathcal{P}$ and $\mathcal{P}_d$ be the distributions of the normalized off-diagonal entries $\sqrt{N}H_{ij}$ and the normalized diagonal entries $\sqrt{N}H_{ii}$, respectively. 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.*
4. *The density function g_d of $\mathcal{P}_d$ satisfies the assumptions 1–3.*

Let $h = -g'/g$ and $h_d = -g_d'/g_d$. For a spiked Wigner matrix M in Definition 2 that satisfies Assumption 1, define a matrix $\widetilde{M}$ by

$$\widetilde{M}_{ij} = \frac{1}{\sqrt{F^H N}} h(\sqrt{N} M_{ij}) \quad (i \neq j), \tag{20}$$

$$\widetilde{M}_{ii} = \sqrt{\frac{w_2}{F_d^H N}} h_d\Big(\sqrt{\frac{N}{w_2}} M_{ii}\Big), \tag{21}$$

where

$$F^H = \int_{-\infty}^{\infty} \frac{g'(w)^2}{g(w)} \mathrm{d}w, \quad F_d^H = \int_{-\infty}^{\infty} \frac{g_d'(w)^2}{g_d(w)} \mathrm{d}w.$$

The transformed matrix $\widetilde{M}$ is not a spiked Wigner matrix anymore. Nevertheless, as we will prove in Theorem 7 in Section 5, the CLT for the LSS of $\widetilde{M}$ holds with the mean $m_{\widetilde{M}}(f)$ and the variance $V_{\widetilde{M}}(f)$. Denote by $m_{\widetilde{M}_0}(f)$ the mean $m_{\widetilde{M}}(f)$ with $\lambda = 0$. Then, as in Section 3, we need to maximize

$$\left| \frac{m_{\widetilde{M}}(f) - m_{\widetilde{M}_0}(f)}{\sqrt{V_{\widetilde{M}}(f)}} \right|.$$

In Theorem 8, we prove that the maximum is attained if and only if $f(x) = C_1 \widetilde{\phi}_\lambda(x) + C_2$ for some constants C_1 and C_2, where

$$\begin{aligned}\widetilde{\phi}_\lambda(x) := \log\left(\frac{1}{1 - \sqrt{\lambda F^H} x + \lambda F^H}\right) \\ + \sqrt{\lambda}\left(\frac{2\sqrt{F_d^H}}{w_2} - \sqrt{F^H}\right) x \\ + \lambda\left(\frac{G^H}{\widetilde{w_4} - 1} - \frac{F^H}{2}\right) x^2.\end{aligned} \tag{22}$$

Thus, denoting by $\widetilde{\mu}_1 \geq \widetilde{\mu}_2 \geq \cdots \geq \widetilde{\mu}_N$ the eigenvalues of $\widetilde{M}$, we define the test statistic $\widetilde{L}_\lambda$ by

$$\begin{aligned}\widetilde{L}_\lambda &:= \sum_{i=1}^N \widetilde{\phi}_\lambda(\widetilde{\mu}_i) - N \int_{-2}^{2} \frac{\sqrt{4 - z^2}}{2\pi} \widetilde{\phi}_\lambda(z)\, \mathrm{d}z \\ &= -\log\det\left((1 + \lambda F^H) I - \sqrt{\lambda F^H} \widetilde{M}\right) + \frac{\lambda F^H}{2} N \\ &\quad + \sqrt{\lambda}\left(\frac{2\sqrt{F_d^H}}{w_2} - \sqrt{F^H}\right) \operatorname{Tr} \widetilde{M} \\ &\quad + \lambda\left(\frac{G^H}{\widetilde{w_4} - 1} - \frac{F^H}{2}\right) (\operatorname{Tr} \widetilde{M}^2 - N),\end{aligned} \tag{23}$$

where

$$G^H = \frac{1}{2F^H} \int_{-\infty}^{\infty} \frac{g'(w)^2 g''(w)}{g(w)^2} \mathrm{d}w,$$

and

$$\widetilde{w_4} = \frac{1}{(F^H)^2} \int_{-\infty}^{\infty} \frac{(g'(w))^4}{(g(w))^3} \mathrm{d}w.$$

The CLT for $\widetilde{L}_\lambda$ holds as follows:

Theorem 3. *Let M be a spiked Wigner matrix in Definition 2 that satisfy Assumption 1. Suppose that $\lambda < \frac{1}{F^H}$. Then,*

$$\widetilde{L}_\lambda \Rightarrow \mathcal{N}(\widetilde{m}_0, \widetilde{V}_0) \qquad \text{if } \lambda = 0$$

and

$$\widetilde{L}_\lambda \Rightarrow \mathcal{N}(\widetilde{m}_+, \widetilde{V}_0) \qquad \text{if } \lambda > 0.$$

The means and the variance of the limiting Gaussian distributions are given by

$$\begin{aligned}\widetilde{m}_0 = -\frac{1}{2}\log(1 - \lambda F^H) \\ + \left(\frac{(w_2 - 1)G^H}{\widetilde{w_4} - 1} - \frac{F^H}{2}\right)\lambda + \frac{\widetilde{w_4} - 3}{4}(\lambda F^H)^2,\end{aligned}$$

$$\begin{aligned}\widetilde{m}_+ = \widetilde{m}_0 - \log(1 - \lambda F^H) \\ + \left(\frac{2F_d^H}{w_2} - F^H\right)\lambda + \left(\frac{(G^H)^2}{\widetilde{w}_4 - 1} - \frac{(F^H)^2}{2}\right)\lambda^2,\end{aligned}$$

and

$$\begin{aligned}\widetilde{V}_0 = -2\log(1 - \lambda F^H) + \left(\frac{4F_d^H}{w_2} - 2F^H\right)\lambda \\ + \left(\frac{2(G^H)^2}{\widetilde{w}_4 - 1} - (F^H)^2\right)\lambda^2.\end{aligned}$$

Theorem 3 is a direct consequence of a general CLT in Theorem 7 in Section 5.

With the entrywise transformation, we modify the hypothesis test as in Algorithm 2, where we compute $\widetilde{L}_\lambda$ and compare it with

$$\begin{aligned}\widetilde{m}_\lambda &:= \frac{\widetilde{m}_0 + \widetilde{m}_+}{2} \\ &= -\log(1 - \lambda F^H) \\ &\quad + \left(\frac{F_d^H}{w_2} - F^H + \frac{(w_2 - 1)G^H}{\widetilde{w_4} - 1}\right)\lambda \\ &\quad + \left(\frac{\widetilde{w_4}}{4} - 1\right)(\lambda F^H)^2 + \frac{(\lambda G^H)^2}{2(\widetilde{w_4} - 1)}.\end{aligned} \tag{24}$$

The parameters w_2 and $\widetilde{w}_4$ can be estimated from the data matrix as in Algorithm 1, and the densities can be estimated by methods such as the kernel density estimation.

Theorem 4. *The error of the test in Algorithm 2,*

$$\operatorname{err}(\lambda) = \mathbb{P}(\widetilde{L}_\lambda > \widetilde{m}_\lambda | \boldsymbol{H}_0) + \mathbb{P}(\widetilde{L}_\lambda \leq \widetilde{m}_\lambda | \boldsymbol{H}_1),$$

Algorithm 2 Hypothesis test with entrywise transformation

Input: data M_{ij}, parameters w_2, w_4, λ, densities g, g_d
$\widetilde{M} \leftarrow$ transformed matrix in Equations (20) and (21)
$\widetilde{L}_\lambda \leftarrow$ test statistic in (23)
$\widetilde{m}_\lambda \leftarrow$ critical value in (24)
if $\widetilde{L}_\lambda \le \widetilde{m}_\lambda$ **then**
  Accept $\boldsymbol{H}_0$
else
  Reject $\boldsymbol{H}_0$
end if

converges to $\operatorname{erfc}(\widetilde{E}_\lambda/4)$*, where*

$$\widetilde{E}_\lambda^2 = \log\left(\frac{1}{1-\lambda F^H}\right) + \left(\frac{2F_d^H}{w_2} - F^H\right)\lambda + \left(\frac{(G^H)^2}{\widetilde{w_4}-1} - \frac{(F^H)^2}{2}\right)\lambda^2.$$

The proof closely follows the proof of Theorem 2, and we omit the detail.

Example 1. Consider the case where the density function of the noise matrix is given by

$$g(x) = g_d(x) = \frac{1}{2\cosh(\pi x/2)} = \frac{1}{e^{\pi x} + e^{-\pi x}}.$$

Sample $W_{ij} = W_{ji}$ from the density g and let $H_{ij} = W_{ij}/\sqrt{N}$. Let $\boldsymbol{x} = (x_1, x_2, \dots, x_N)$ where $\sqrt{N}x_i$'s are i.i.d. Rademacher random variable. Let the data matrix $M = \sqrt{\lambda}\boldsymbol{x}\boldsymbol{x}^T + H$. The parameters are $w_2 = 1$, $w_4 = 5$.

In the test proposed in Section 3, we compute

$$L_\lambda = -\log\det\left((1+\lambda)I - \sqrt{\lambda}M\right) + \frac{\lambda N}{2} + \sqrt{\lambda}\operatorname{Tr} M - \frac{\lambda}{4}(\operatorname{Tr} M^2 - N), \tag{25}$$

and accept $\boldsymbol{H}_0$ if $L_\lambda \le -\log(1-\lambda) + \frac{3\lambda^2}{8}$ and reject $\boldsymbol{H}_0$ otherwise. The limiting error of the test is

$$\operatorname{erfc}\left(\frac{1}{4}\sqrt{-\log(1-\lambda) + \lambda - \frac{\lambda^2}{4}}\right). \tag{26}$$

We can further improve the test by introducing the entrywise transformation given by

$$h(x) = -\frac{g'(x)}{g(x)} = \frac{\pi}{2}\tanh\frac{\pi x}{2}.$$

The Fisher information F^H is $\frac{\pi^2}{8}$, which is strictly larger than 1. We first construct a pre-transformed matrix $\widetilde{M}$ by

$$\widetilde{M}_{ij} = \frac{2\sqrt{2}}{\pi\sqrt{N}}h(\sqrt{N}M_{ij}) = \sqrt{\frac{2}{N}}\tanh\left(\frac{\pi\sqrt{N}}{2}M_{ij}\right).$$

If $\lambda > \frac{1}{F^H} = \frac{8}{\pi^2}$, we can use PCA to reliably detect the signal. If $\lambda < \frac{8}{\pi^2}$, we compute the test statistic

$$\widetilde{L}_\lambda = -\log\det\left((1+\frac{\pi^2\lambda}{8})I - \sqrt{\frac{\pi^2\lambda}{8}}\widetilde{M}\right) + \frac{\pi^2\lambda N}{16} + \frac{\pi\sqrt{\lambda}}{2\sqrt{2}}\operatorname{Tr}\widetilde{M} + \frac{3\pi^2\lambda}{16}(\operatorname{Tr}\widetilde{M}^2 - N).$$

(Here, $G^H = F^H = \frac{\pi^2}{8}$ and $\widetilde{w}_4 = \frac{3}{2}$.) We accept $\boldsymbol{H}_0$ if

$$\widetilde{L}_\lambda \le -\log\left(1 - \frac{\pi^2\lambda}{8}\right) + \frac{3\pi^4\lambda^2}{512}$$

and reject $\boldsymbol{H}_0$ otherwise. The limiting error with entrywise transformation is

$$\operatorname{erfc}\left(\frac{1}{4}\sqrt{-\log(1-\frac{\pi^2\lambda}{8}) + \frac{\pi^2\lambda}{8} + \frac{3\pi^4\lambda^2}{128}}\right). \tag{27}$$

Since $\operatorname{erfc}(z)$ is a decreasing function of z and $\frac{\pi^2}{8} > 1$, it is direct to see that the limiting error in (27) is strictly less than the limiting error in (26) as illustrated in Figure 2.

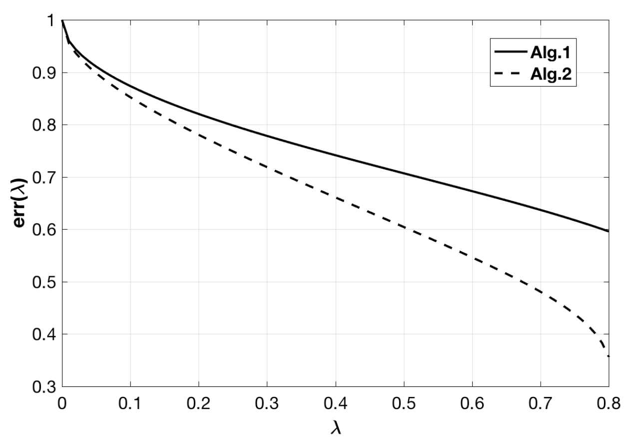

Figure 2. The limiting errors (26) of Algorithm 1 (solid) and (27) of Algorithm 2 (dashed), respectively, for Example 1.

5. Central Limit Theorems

In this section, we collect general CLTs for the LSS. The mean and the variance will be written in terms of Chebyshev polynomials (of the first kind) for which we use the following definition.

Definition 3 (Chebyshev polynomial)**.** *The n-th Chebyshev polynomial T_n is a degree n polynomial defined by the orthogonality condition*

$$\int_{-1}^{1} T_m(x)T_n(x)\frac{\mathrm{d}x}{\sqrt{1-x^2}} = \begin{cases} 0 & \textit{if } m \ne n, \\ \pi & \textit{if } m = n = 0, \\ \frac{\pi}{2} & \textit{if } m = n \ne 0. \end{cases}$$

Our first result in this section is a general CLT for the LSS.

Theorem 5. *Assume the conditions in Theorem 1. For any function f analytic on an open interval containing $[-2, 2]$,*

$$\left(\sum_{i=1}^{N} f(\mu_i) - N\int_{-2}^{2} \frac{\sqrt{4-z^2}}{2\pi} f(z)\,\mathrm{d}z\right) \Rightarrow \mathcal{N}(m_M(f), V_M(f))\,.$$

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

$$m_M(f) = \frac{1}{4}\left(f(2)+f(-2)\right) - \frac{1}{2}\tau_0(f) + (w_2-2)\tau_2(f) + (w_4-3)\tau_4(f) + \sum_{\ell=1}^{\infty}\sqrt{\lambda^\ell}\tau_\ell(f)$$

and

$$V_M(f) = (w_2-2)\tau_1(f)^2 + 2(w_4-3)\tau_2(f)^2 + 2\sum_{\ell=1}^{\infty}\ell\tau_\ell(f)^2,$$

where

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

with the ℓ-th Chebyshev polynomial T_ℓ.

Our second result classifies all functions that are optimal for the hypothesis test.

Theorem 6. *Assume the conditions in Theorem 5. If $w_2 > 0$ and $w_4 > 1$, then*

$$\left|\frac{m_M(f)-m_H(f)}{\sqrt{V_M(f)}}\right| \le \left|\frac{m_+ - m_0}{\sqrt{V_0}}\right|. \tag{28}$$

The equality holds if and only if $f = C_1\phi_\lambda + C_2$ for some constants C_1 and C_2 with the function ϕ_λ defined in (7).

The function of the form ϕ_λ in (7) was considered by Banerjee and Ma for hypothesis testing in stochastic block models; see Remark 3.3 in (Banerjee & Ma, 2017). Instead of using polynomial approximation of ϕ_λ as in (Banerjee & Ma, 2017), we use ϕ_λ itself since it is analytic for any x in an open interval $(-\sqrt{\lambda}-\frac{1}{\sqrt{\lambda}}, \sqrt{\lambda}+\frac{1}{\sqrt{\lambda}})$, which contains $[-2, 2]$. In the signal detection test we consider, if there is an eigenvalue outside the interval $(-\sqrt{\lambda}-\frac{1}{\sqrt{\lambda}}, \sqrt{\lambda}+\frac{1}{\sqrt{\lambda}})$, it implies that the signal is present with high probability.

Our result for the pre-transformed CLT is the following theorem:

Theorem 7. *Assume the conditions in Theorem 3. For any function f analytic on an open interval containing $[-2, 2]$,*

$$\left(\sum_{i=1}^{N} f(\widetilde{\mu}_i) - N\int_{-2}^{2} \frac{\sqrt{4-z^2}}{2\pi} f(z)\,\mathrm{d}z\right) \Rightarrow \mathcal{N}(m_{\widetilde{M}}(f), V_{\widetilde{M}}(f))\,.$$

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

$$m_{\widetilde{M}}(f) = \frac{1}{4}\left(f(2)+f(-2)\right) - \frac{1}{2}\tau_0(f) + \sqrt{\lambda F_d^H}\tau_1(f) + (w_2 - 2 + \lambda G^H)\tau_2(f) + (\widetilde{w_4}-3)\tau_4(f) + \sum_{\ell=3}^{\infty}\sqrt{(\lambda F^H)^\ell}\tau_\ell(f) \tag{29}$$

and

$$V_{\widetilde{M}}(f) = (w_2-2)\tau_1(f)^2 + 2(\widetilde{w_4}-3)\tau_2(f)^2 + 2\sum_{\ell=1}^{\infty}\ell\tau_\ell(f)^2.$$

Note that $V_{\widetilde{M}}(f)$ does not depend on the existence of the spike. Let $m_{\widetilde{M}_0}(f)$ be $m_{\widetilde{M}}(f)$ in (29) with $\lambda = 0$. For the transformed matrix $\widetilde{M}$, we have the following result that corresponds to Theorem 6.

Theorem 8. *Assume the conditions in Theorem 7. Then*

$$\left|\frac{m_{\widetilde{M}}(f)-m_{\widetilde{M}_0}(f)}{\sqrt{V_{\widetilde{M}}(f)}}\right| \le \sqrt{\frac{\widetilde{m}_+ - \widetilde{m}_0}{\widetilde{V}_0}}. \tag{30}$$

The equality holds if and only if $f(x) = C_1\widetilde{\phi}_\lambda(x) + C_2$ for some constants C_1 and C_2 with $\widetilde{\phi}_\lambda$ defined in (22).

6. Conclusion and Future Works

In this paper, we proposed a hypothesis test for a signal detection problem in a rank-one spiked Wigner model. Based on the central limit theorem for the linear spectral statistics of the data matrix, we established a test statistic that does not require any prior information on the signal. The test and its error depends on the noise matrix only through the variance of the diagonal entries and the fourth moment of the off-diagonal entries. The error of the proposed test is the lowest among all tests based on the linear spectral statistics, and it also matches the error of the likelihood ratio test if the noise is Gaussian. When the density of the noise is known, we further improve the test by adapting the entrywise transformation introduced in (Perry et al., 2018).

An interesting future research direction is to extend the test to the case with a spike of higher rank. We believe that it is possible to prove the central limit theorem for the linear statistics even when the rank of the signal is higher, and our test can be naturally extended to the model. We also hope to generalize our results to the data matrix with non-Wigner noise, where the variances of off-diagonal entries of the noise matrix are not identical, including (sparse) stochastic block models.

Acknowledgements

The authors would like to thank anonymous reviewers for their suggestions and feedback. The work of H. W. Chung was partially supported by National Research Foundation of Korea under grant number 2017R1E1A1A01076340 and by the Ministry of Science and ICT, Korea, under an ITRC Program, IITP-2019-2018-0-01402. The work of J. O. Lee was partially supported by the Samsung Science and Technology Foundation project number SSTF-BA1402-04.

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 Yao, J. On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli*, 11(6): 1059–1092, 2005.

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. Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model with ferromagnetic interaction. *Ann. Henri Poincaré*, 18(6):1867–1917, 2017.

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.

Banerjee, D. and Ma, Z. Optimal hypothesis testing for stochastic block models with growing degrees. *arXiv:1705.05305*, 2017.

Barbier, j., Dia, M., Macris, N., Krzakala, F., Lesieur, T., and Zdeborová, L. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems 29*, pp. 424–432. 2016.

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.

Butucea, C., Ingster, Y. I., et al. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli*, 19(5B):2652–2688, 2013.

Capitaine, M., Donati-Martin, C., and Féral, D. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Ann. Probab.*, 37(1):1–47, 2009.

El Alaoui, A., Krzakala, F., and Jordan, M. I. Fundamental limits of detection in the spiked Wigner model. *arXiv:1806.09588*, 2018.

Erdős, L., Yau, H.-T., and Yin, J. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.*, 229(3):1435–1515, 2012.

Féral, D. and Péché, S. The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.*, 272(1):185–228, 2007.

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

Johnstone, I. M. and Lu, A. Y. On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.*, 104(486):682–693, 2009.

Lelarge, M. and Miolane, L. Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields*, 173(3-4):859–929, 2019.

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.

Nadler, B. Finite sample approximation results for principal component analysis: a matrix perturbation approach. *Ann. Statist.*, 36(6):2791–2817, 2008.

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

Paul, D. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica*, 17 (4):1617–1642, 2007.

Péché, S. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields*, 134(1):127–173, 2006.

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.

Tao, T. and Vu, V. Random matrices: universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.*, 298(2):549–572, 2010.

Tracy, C. A. and Widom, H. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.

Tracy, C. A. and Widom, H. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177(3):727–754, 1996.