Proceedings of Machine Learning Research vol ???:1–34, 2025

Minimax Detection of the Number of Spikes in Large Wigner Matrices

Madhav Sankaranarayanan MADHAV_SANKARANARAYANAN@G.HARVARD.EDU
Harvard T.H. Chan School of Public Health, Boston, MA, USA

Soumendu Sundar Mukherjee SSMUKHERJEE@ISICAL.AC.IN
Indian Statistical Institute, Kolkata, India

Rajarshi Mukherjee RMUKHERJ@HSPH.HARVARD.EDU
Harvard T.H. Chan School of Public Health, Boston, MA, USA

Abstract

We investigate the problem of minimax optimal detection of the number of spikes in lower rank signal plus large Wigner models. In this regard, we categorize sharp conditions for asymptotic power in testing for the number of spikes being k_0 versus k_1 in terms of k_1, k_0 and the spectral gap between the first k_0 and last $k_1 - k_0$ spiked eigenvalues. Additionally, when spikes are bounded, we elucidate a novel optimal test in the subcritical regime under a Gaussian Orthogonal Ensemble. For the general hypothesis testing problem, we compare the performance of some ubiquitous tests in spike detection problems. We also elucidate an algorithm for estimating the number of spikes in outlying spiked setups.

Keywords: Spiked Wigner matrices, Gaussian Orthogonal Ensembles, Likelihood ratio, Bipartite signed cycles, Sequential hypothesis testing

1. Introduction

Statistical inference in high-dimensional spiked matrix models (Johnstone, 2006; Paul, 2007; Péché, 2006; Bai and Yao, 2008; Baik et al., 2004; Féral and Péché, 2007; Capitaine et al., 2009; Benaych-Georges and Nadakuditi, 2011; Baik and Silverstein, 2006; Ke, 2016) has witnessed continued interest from quantitative scientists owing to its broad range of applications across many different fields that include but are not limited to wireless communications (Telatar, 1999), genetic association studies (Patterson et al., 2006; Price et al., 2006), and financial portfolio optimization (Sharifi et al., 2004; Bai et al., 2009). Whereas estimation and detection of spikes, i.e., outlying singular vectors for large matrix models, have received considerable attention from the research community (Onatski et al., 2014; El Alaoui and Jordan, 2018; El Alaoui et al., 2020; Onatski et al., 2013, 2014; Lelarge and Miolane, 2017; Dobriban, 2017; Banerjee and Ma, 2018; Dia et al., 2016; Ke, 2016; Mukherjee, 2023; Kunisky, 2024; Chung et al., 2022; Jung et al., 2021; Dudeja et al., 2024; Pak et al., 2024; El Alaoui et al., 2020), an analogous information theoretic analysis for estimating and testing the number of outlying singular vectors or signals is rare – and only a few papers (Jung et al., 2020, 2023; Chung and Lee, 2019) in recent times have made significant progress in this regard. In this paper, we, therefore, formalize a minimax framework for testing for the number of outlying eigenvalues in the spiked Wigner model to establish matching lower and upper bounds for both fixed and diverging numbers of spikes – as well as discuss the main differences of our results compared to the aforementioned literature. The main results of this paper pertain to sharp analysis of the testing problem up to precise multiplicative constants in the fixed number of spikes regime and rate optimal analysis of the minimax problem in the diverging number of spikes regime.

The rest of the paper is organized as follows. In Section 1.1, we collect some notations and terminology that will be used, followed by the mathematical formalism and the minimax framework in Section 1.2. Subsequently, in Section 2, we collect the information-theoretic lower bounds, followed by an analysis of optimal and near-optimal testing procedures in Section 3. Next we present the proof ideas in Section 4 before collecting the proofs in the Appendix. Finally, we conclude with discussions of future directions in Section 6.

1.1. Notation and terminology

We use the following notation and terminology in the manuscript. We will denote e_i as the ith canonical basis vector in $\mathbb{R}^n$. $\mathrm{diag}(u)$ will be used to represent a diagonal $n \times n$ matrix with diagonal elements $u \in \mathbb{R}^n$. For matrices, $||\cdot||_{\max}$ refers to the element-wise maximum value, $||\cdot||_F^2$ refers to the Frobenius norm and $||\cdot||_{op}$ refers to the operator norm. Moreover, $\mathbb{S}^n$ will denote the set of $n \times n$ orthogonal matrices over $\mathbb{R}$ and $\mathbb{S}^{p\times\kappa}$ will be used to represent the Stiefel manifold consisting of all κ-frames in $\mathbb{R}^p$. We will call an $n \times n$ matrix Wigner if it can be written as $\frac{Y+Y^\top}{\sqrt{2}}$ with Y being an upper triangular matrix with i.i.d. elements from some distribution F having mean 0, variance 1, and eight finite moments. A scaled Wigner matrix is simply a matrix of the form $\frac{Y+Y^T}{\sqrt{2n}}$. A specific scaled Wigner matrix, $\mathrm{GOE}(n)$ will stand for Gaussian Orthogonal Ensemble of size $n \times n$ (scaled by $\sqrt{2n}$), that is a Wigner matrix with normal entries, scaled by $\sqrt{2n}$. We will omit the (n) when clear from context. We use $\preceq$ to denote the partial Loewner ordering, that is if $A \preceq B$, then $B - A$ is positive semi-definite. We also use asymptotic notations such as $o(\cdot), \mathcal{O}(\cdot), \Theta(\cdot)$ and $\Omega(\cdot)$, and probabilistic asymptotic notations, such as $\mathcal{O}_{\mathbb{P}}(\cdot)$ and $o_{\mathbb{P}}(\cdot)$, where these notations carry their usual meaning.

Additionally, we invoke the notion of contiguity and absolute continuity in our proofs. Consider two sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ defined on the σ-fields $(\Omega_n, \mathcal{F}_n)$, we say that $\mathbb{Q}_n$ is contiguous with respect to $\mathbb{P}_n$ if for any event sequence A_n, $\mathbb{P}_n(A_n) \to 0$ implies $\mathbb{Q}_n(A_n) \to 0$. We say that the two sequences are asymptotically mutually contiguous if $\mathbb{Q}_n$ is contiguous with respect to $\mathbb{P}_n$ and $\mathbb{P}_n$ is contiguous with respect to $\mathbb{Q}_n$. The notion of contiguity is necessary from a testing perspective if we consider the two probability measures to be the distributions of the test statistic under the null and alternate hypothesis. This is closely related to the concept of absolute continuity with respect to measures, which will also be invoked in the proofs of the main results of the paper.

1.2. Setup and mathematical formalization

We formalize testing for the number of spikes in a Wigner matrix as the following hypothesis testing problem. Specifically we consider the spiked Wigner model defined for an observed a $n \times n$ symmetric matrix X as

$$X = U\Lambda U^T + W,$$

where $\Lambda = \mathrm{diag}((\tilde{\lambda}_i)), \tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \ldots \geq \tilde{\lambda}_n \geq 0$, U is an orthogonal matrix belonging to $\mathbb{S}^n$, and W is a $n \times n$ real (scaled) Wigner random matrix (Wigner, 1967) independent of U with elements drawn from distribution F as introduced in Section 1.1. Indeed a special case of this model is the spiked GOE where W is drawn from a Gaussian Orthogonal Ensemble. In this model, we formalize testing for a number of spiked eigenvalues being k_0 versus $k_1 > k_0$ as follows. Suppose we have primary spikes $\tilde{\lambda}_1, \dots, \tilde{\lambda}_{k_0} \geq \lambda_0$ and secondary spikes $\lambda_1 \geq \tilde{\lambda}_{k_0+1}, \dots, \tilde{\lambda}_{k_1} > 0$, whose presence we aim to detect, with $\tilde{\lambda}_{k_1+1}, \dots, \tilde{\lambda}_n = 0$ and $\beta \geq \lambda_0 > \lambda_1 > 0$ for some fixed $\beta < \infty$. The main purpose of this paper is to derive minimax optimal tests for

$$H_0 : \ \lambda_1 = 0 \ \text{ vs. } \ H_1 : \ \lambda_1 > \lambda, \tag{1}$$

where $\lambda > 0$. Note that here we consider unidirectional spikes for simplicity of exposition and we note that the lower bounds do not really depend on this assumption and the upper bounds can be easily modified by the literature cited in the proofs and exposition. We define a test as a measurable function of X that returns 1 when the null hypothesis is rejected and 0 if there is a failure to reject the null. We denote $\mathcal{F}_{\mathbb{S}^n}$ to be the set of distributions over $\mathbb{S}_n$, which allows for Dirac masses implying fixed U. For a given test $T_n(X)$, we can define the risk associated with testing this hypothesis as follows:

$$\mathrm{Risk^W}_{n,\lambda}(T_n) := \sup_F \sup_{f \in \mathcal{F}_{\mathbb{S}^n}} \sup_{\beta > \lambda_0 > \lambda_1 = 0} \mathbb{P}_{H_0, U \sim f}(T_n = 1) + \sup_F \sup_{f \in \mathcal{F}_{\mathbb{S}^n}} \sup_{\beta > \lambda_0 > \lambda_1 > \lambda} \mathbb{P}_{H_1, U \sim f}(T_n = 0).$$

For the GOE case, we remove the supremum over the Wigner distributions, and use a simplified notation for the risk function.

$$\mathrm{Risk}_{n,\lambda}(T_n) := \sup_{f \in \mathcal{F}_{\mathbb{S}^n}} \sup_{\beta > \lambda_0 > \lambda_1 = 0} \mathbb{P}_{H_0, U \sim f}(T_n = 1) + \sup_{f \in \mathcal{F}_{\mathbb{S}^n}} \sup_{\beta > \lambda_0 > \lambda_1 > \lambda} \mathbb{P}_{H_1, U \sim f}(T_n = 0).$$

The second nested supremum is over the values of $\tilde{\lambda}_i$ subject to the given inequality constraints. We say that a sequence of tests T_n is asymptotically powerful if $\limsup_{n \to \infty} \mathrm{Risk}_{n,\lambda}(T_n) = 0$ and is asymptotically powerless if $\liminf_{n \to \infty} \mathrm{Risk}_{n,\lambda}(T_n) = 1$. Additionally, we call a sequence of tests T_n asymptotically not powerful/nontrivial power if $0 < \liminf_{n \to \infty} \mathrm{Risk}_{n,\lambda}(T_n) < \limsup_{n \to \infty} \mathrm{Risk}_{n,\lambda}(T_n) < 1$. The ability to perform this hypothesis testing problem relies on the interplay between the strength of the secondary spikes, codified by λ_1, and the sparsity of the secondary spikes, i.e., spectral gap codified by $k_1 - k_0$. In this framework, Section 2 focuses on analyzing the regimes of $(k_1 - k_0)$ and λ_1 where no tests are asymptotically powerful as well deriving the precise optimal power function for bounded k_1 under the special case of $\mathrm{GOE}(n)$ distributed W and Haar distributed U. Establishing these testing limits will give us a proper understanding of the extent of testing possible for the number of spikes in a matrix model.

Subsequently, in Section 3, we focus on regimes of the problem where one can potentially find asymptotically powerful tests and compare efficacy of practically applicable tests. We characterize the asymptotic power of these tests for testing the existence of secondary spikes (which we subsequently refer to as simply "spikes"). The application of these tests in sequentially defined testing problems allows us to consistently estimate the number of spikes in a spiked model, which will be described in the Appendix.

2. Lower Bounds: Regimes where Asymptotically Powerful Tests Don't Exist

We divide the study of lower bounds of the problem according to k_1 being bounded or not. When k_1 is bounded, we furnish a lower bound by considering the special case where W follows a GOE and U follows a uniform distribution on $\mathbb{S}^n$. This allows us to derive the precise behavior of the risk below the information-theoretic threshold. Indeed, substantial literature have derived deep results for the testing problem when $k_0 = 0$ and/or k_1 is bounded – see e.g. (Onatski et al., 2014; El Alaoui and Jordan, 2018; El Alaoui et al., 2020; Onatski et al., 2013, 2014; Lelarge and Miolane, 2017; Dobriban, 2017; Banerjee and Ma, 2018; Dia et al., 2016; Ke, 2016; Mukherjee, 2023; Kunisky, 2024; Chung et al., 2022; Jung et al., 2021; Dudeja et al., 2024; Pak et al., 2024; El Alaoui et al., 2020) and references therein. In this regard, there is a subtle difference in the structure of the testing problem in Jung et al. (2020, 2023); Chung and Lee (2019) and our work. Additionally, we derive another sequence of tests based on signed cycles that is easily implementable and can obtain sharp asymptotic error behavior with precise error functions provided in the theorem below and its proof. Subsequently, we go beyond the regimes of a fixed number of spikes provided in Jung et al. (2020, 2023); Chung and Lee (2019) and derive the information-theoretic lower bounds to determine regimes of k_1, k_0, λ_1 where no asymptotically powerful tests exist. In this regard, a parallel paper of Ke (2016) derives information theoretic lower bound of the PCA version of the problem when k_1 scales polynomial in the dimension. Our results, however, accommodates general regimes of unbounded k_1, k_0.

For the spiked GOE setup elucidated above, we are interested in finding the conditions under which distinguishing secondary spikes is statistically impossible. The results in the following section arise from the analysis of the power of the likelihood ratio test. Based on our notations from before, k_1 represents the total number of spikes and k_0 represents the number of primary (larger) spikes. Each of the primary spikes has a value of λ_0, and each secondary spike has a value of λ_1. In order to introduce the relevant theorems, we need to define some preliminary objects of interest. We denote the density of X under the null as f_0 and under the alternate as f_1. The likelihood ratio L_n is defined as $L_n = \frac{f_1}{f_0}$.

We define lower diagonal bipartite signed cycles of length 2ℓ, denoted by $B_{n,\ell}$, as

$$B_{n,\ell} = 2^{\ell/2} \sum_{i_0, j_0, \ldots, i_{\ell-1}, j_{\ell-1}} X_{i_0 j_0} X_{i_1 j_0} X_{i_1 j_1} \ldots X_{i_{\ell-1} j_{\ell-1}} X_{i_0 j_{\ell-1}},$$

where $i_0, i_1, \ldots, i_{\ell-1} \in \{1, \ldots, n\}$ all distinct and $j_0, j_1, \ldots, j_{\ell-1} \in \{1, \ldots, n\}$ all distinct. Let $\mu(\ell)$ denote the expectation of $B_{n,\ell}$ and $v(\ell)$ denote the variance of $B_{n,\ell}$.

If the number of these secondary spikes is bounded, we observe the following result involving the region of impossibility.

Theorem 1 *Consider the testing problem given (1) with k_1 bounded.*

(a) If $\lambda < 1$, then no test is asymptotically powerful.

(b) In this subcritical regime, the optimal test is given by the likelihood ratio test $\mathbb{1}_{\log(L_n) \geq 0}$, with the following risk function

$$\limsup_{n \to \infty} \mathrm{Risk}_{n,\lambda}(\log(L_n)) = 2\left(1 - \Phi\left(\tilde{\mu}/\sqrt{\tilde{v}}\right)\right) \leq \mathrm{erfc}\left(\frac{(k_1 - k_0)}{4}\sqrt{\log\left(\frac{1}{1-\lambda}\right)}\right),$$

where

$$\tilde{\mu} = \sum_{\ell=1}^{\infty} \frac{\mu(\ell)^2}{2\ell}, \tilde{v} = \sum_{\ell=1}^{\infty} \frac{\mu(\ell)^2 v(\ell)}{4\ell^2}$$

and erfc *is the error function defined by* $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$

The interpretation of this result is that for a bounded number of spikes, if the secondary spike strength is smaller than an exact cutoff, the distributions under the null and alternate hypotheses in (1) are mutually contiguous. Additionally, the likelihood ratio test has non-zero power for non-trivial values of λ_1 asymptotically. This represents an exact lower bound of testing for this setup, and is the relevant lower bound for a sequential testing framework for establishing the number of spikes. It is to be noted that Theorem 3.2 of Chung and Lee (2019) obtains the same asymptotic risk of the likelihood ratio test, despite the underlying testing problem being different (Chung and Lee (2019) has a simple null with $\lambda_0 = \lambda_1 = \lambda$, while our work focuses on a composite null). The case of a constant signal-to-noise ratio λ allows for the use of the monotonicity of linear spectral statistics to prove this lower bound. In our case, the additional structure does not change the lower bound, but influences the proof techniques we can employ to attain this lower bound. We elucidate the skeleton of the proof of this theorem in Section 4, which uses the construction of relevant combinatorial objects known as "bipartite signed cycles", which were introduced as a mechanism for proving contiguity in Banerjee and Ma (2018). The enumeration of these objects is a key insight in this proof, and requires novel combinatorial insights. For the bounded spike sparsity case, we implement simulations to show the power of this bipartite signed cycle test in Section 5.

In our next result we consider the case of diverging number of spikes and go beyond the results in Jung et al. (2020, 2023); Chung and Lee (2019). When the number of spikes goes to infinity, we have the following result on the information-theoretic lower bound.

Theorem 2 *Consider the testing problem given in (1) with k_1 unbounded. If $\lambda_1(k_1 - k_0) = \mathcal{O}(1)$, then no tests are asymptotically powerful.*

Remark

1. *In the setup given in Theorem 2, if $\lambda_1 (k_1 - k_0) = o(1)$, then all tests are asymptotically powerless.*

2. *This result is parallel to the result in spiked PCA in Ke (2016) where the number of spikes was characterized as a polynomial of the dimension of the problem and only considered the case of $k_0 = 0$. Our proof generalizes this for general k_1, k_0 and without the polynomial scaling.*

Unlike Theorem 1, this theorem shows that the region of impossibility depends on the "total secondary signal", or the product of the secondary spike strength and the number of secondary spikes – and hence allows for decaying spectral gap λ_1. If this signal is finite, then the distributions under the null and alternate hypotheses in (1) are mutually contiguous. We describe the proof of this theorem in Section 4, which requires a careful analysis of the likelihood ratio.

3. Upper Bounds: Regimes where Asymptotically Powerful Tests Exist

Recall the definition of the Wigner matrix $W = Y + Y^T$, with upper diagonal Y with each element drawn from F. We will study the performance of various tests under the hypothesis testing framework of (1). We denote the ordered eigenvalues of X as $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_n$.

The tests we will be considering are:

(i) $\mathcal{T}_1$ **: Trace test**, which rejects H_0 if $\mathrm{Tr}(X)$ is large

(ii) $\mathcal{T}_2$ **: Partial trace test**, which rejects H_0, if the k-th order partial trace $\mathrm{PT}_k(X) := \sum_{j=1}^{k} \eta_j$ is large, for some fixed $k > k_1$

(iii) $\mathcal{T}_3$: **Largest noncritical eigenvalue test**, which rejects H_0 is the largest non-critical eigenvalue η_{k_0+1} is large

These tests are ubiquitous for testing problems related to the spectra of random matrices, and are practically implementable.

Theorem 3 *Suppose W is a Wigner matrix with F having subexponential tails. Then*

- *$\mathcal{T}_1$ is asymptotically powerful if and only if $(k_1 - k_0)\lambda_1 \to \infty$.*
- *In the setting of fixed k_1 and $\lambda_1 > 1$, $\mathcal{T}_2$ is asymptotically powerful if*

$$(k_1 - k_0)\frac{(\lambda_1 - 1)^2}{\lambda_1} - k_0\left(\beta + \frac{1}{\beta} - \lambda_0 - \frac{1}{\lambda_0}\right) > 0,$$

- *In the setting of fixed k_1 and $\lambda_1 > 1$, $\mathcal{T}_3$ is asymptotically powerful.*

The proof of Theorem 3 is given in the Appendix, where asymptotic expressions for the power functions of these tests are obtained. In the supercritical regime, all three tests are asymptotically powerful, but the small sample power functions are not directly known. We compare these tests in the supercritical regime through simulations in Section 5.

4. Ideas of Proofs of Main Results

4.1. Proof sketch of Theorem 1

We have the setup from (1) with $k_0 = 0$

$$H_0 : \; X = W \text{ vs. } H_1 : \; X = U\Lambda U^T + W,$$

where $\Lambda = \mathrm{diag}(\lambda_1 \sum_{i=1}^{k_1} e_i)$, $U \sim \mathrm{Unif}(\mathbb{S}^n)$, $W \sim \mathrm{GOE}(n)$, W, U are independent and k_1 bounded.

The structure of the proof, akin to Banerjee and Ma (2018), is as follows:

1. Construct random variables that are "asymptotically sufficient" for the likelihood ratio.
2. Use these variables in the Second Moment method to prove contiguity and derive asymptotic normality.
3. Calculate power for optimal test in subcritical regime.

The ability to construct these random variables arises from the following proposition, a formulation of the second moment method (Proposition 1 in Banerjee and Ma (2018), which is a Gaussian form of Theorem 1 from Janson (1995)).

Proposition 1 (Banerjee and Ma (2018, Proposition 1)) *Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two sequences of probability measures such that for each n, both are defined on the common σ-algebra $(\Omega_n, \mathcal{F}_n)$. Suppose that for each $i \geq 1$, $W_{n,i}$ are random variables defined on $(\Omega_n, \mathcal{F}_n)$. The sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous if the following conditions hold simultaneously:*

(i) $\mathbb{Q}_n$ *is absolutely continuous with respect to* $\mathbb{P}_n$ *for each* n

(ii) *For any fixed* $\ell \geq 1$, *one has* $(W_{n,1}, \ldots, W_{n,\ell})|\mathbb{P}_n \xrightarrow{d} (Z_1, \ldots, Z_\ell)$ *and* $(W_{n,1}, \ldots, W_{n,\ell})|\mathbb{Q}_n \xrightarrow{d} (Z_1', \ldots, Z_\ell')$

(iii) $Z_i \sim N(0, \sigma_i^2)$ *and* $Z_i' \sim N(\mu_i, \sigma_i^2)$ *are sequences of independent random variables*

(iv) *The likelihood ratio statistic* $Y_n = \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$ *satisfies*

$$\limsup_{n\to\infty} \mathbb{E}_{\mathbb{P}_n}[Y_n^2] \leq \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}\right\} < \infty$$

In addition, under these four conditions, we have that under $\mathbb{P}_n$*,*

$$Y_n \xrightarrow{d} \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i Z_i - \frac{1}{2}\mu_i^2}{\sigma_i^2}\right\}$$

Furthermore, given any $\epsilon, \delta > 0$ *there exists a natural number* $K = K(\delta, \epsilon)$ *such that for any sequence* n_l *there is a further subsequence* n_{l_m} *such that*

$$\limsup_{n\to\infty} \mathbb{P}_{n_{l_m}}\left(\left|\log(Y_{n_{l_m}}) - \sum_{\ell=1}^{K} \frac{2\mu_\ell W_{n_{l_m},\ell} - \mu_\ell^2}{2\sigma_\ell^2}\right| \geq \epsilon\right) \leq \delta$$

The random variables $W_{n,i}$ chosen in our formulation are bipartite signed cycles, as given in Section 2. We provide the full set of details on the rationale behind this construction, and the deviation from the bipartite signed cycles presented in Banerjee and Ma (2018) in the Appendix. We define the lower diagonal bipartite signed cycle of length 2ℓ are defined as

$$B_{n,\ell} = 2^{\ell/2} \sum_{\substack{i_0, j_0, \ldots, i_{\ell-1}, j_{\ell-1} \\ j_a \leq i_a \ \forall a \in \{1, \ldots, \ell-1\}}} X_{i_0 j_0} X_{i_1 j_0} X_{i_1 j_1} \ldots X_{i_{\ell-1} j_{\ell-1}} X_{i_0 j_{\ell-1}},$$

where $i_0, i_1, \ldots, i_{\ell-1} \in \{1, \ldots, n\}$ all distinct and $j_0, j_1, \ldots, j_{\ell-1} \in \{1, \ldots, n\}$ all distinct. We use the same notation as Banerjee and Ma (2018), but make the distinction that all the points on these bipartite signed cycles lie in the lower diagonal of our matrix. These random variables exploit the structure of the matrix, and are shown to abide by the stipulations of Proposition 3. The distribution of these objects is codified in the following proposition.

Proposition 2 *Consider the testing problem (1) with* $k_0 = 0$. *Then for any fixed integer* $l > 0$, *there exists a* $f(\ell)$ *with* $f(1) = 1$, $f(\ell)$ *monotonously increases for* $\ell \geq 2$ *and* $f(\ell) \to 1$ *as* $\ell \to \infty$ *such that the following results hold:*

(i) *Under* H_0, *when* $1 \leq \ell_1 < \ldots < \ell_l = o(\sqrt{\log n})$,

$$\left(\frac{B_{n,\ell_1} - \frac{n+3}{\sqrt{2}}\mathbb{1}_{\ell_1=1}}{\sqrt{2\ell_1 f(\ell_1)}}, \cdots, \frac{B_{n,\ell_l}}{\sqrt{2\ell_l f(\ell_l)}}\right) \xrightarrow{d} N_l(0, I_l).$$

(ii) Under H_1, when $1 \leq \ell_1 < \ldots < \ell_l = o(\sqrt{\log n})$,

$$\left(\frac{B_{n,\ell_1} - \frac{n+3}{\sqrt{2}} \mathbb{1}_{\ell_1 = 1} - k\lambda^{2\ell_1} f(\ell_1)}{\sqrt{2\ell_1 f(\ell_1)}}, \cdots, \frac{B_{n,\ell_l} - k\lambda^{2\ell_l} f(\ell_l)}{\sqrt{2\ell_l f(\ell_l)}} \right) \xrightarrow{d} N_l(0, I_l).$$

The proof of this proposition is in the Appendix, and for now, utilize this proposition to prove Theorem 1 in the case where $k_0 = 0$.

The density of our data matrix X is given from the density of a Gaussian Orthogonal Ensemble, and we can use this to construct the likelihood ratio. We define σ-fields $\mathcal{X}_n = \sigma(\{X_{i*}\}_{i=1}^n)$ and $\mathcal{U}_n = \sigma(\{U_{i*}\}_{i=1}^n)$. In terms of X, the conditional likelihood ratio on U is given by

$$L_n^{\mathcal{U}} := \exp\left\{ \frac{n}{2} \sum_{i=1}^n \sum_{j=1}^n \left(X_{i,j} M_{i,j} - \frac{1}{2} M_{i,j}^2 \right) \right\}$$

where $M_{i,j} = \sum_{l=1}^{k_1} \lambda_1 U_{i,l} U_{j,l}$. The probability measures for the null and alternate hypotheses in our analysis are given by $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ respectively.

We can see that our likelihood ratio is given as $L_n = \mathbb{E}[L_n^{\mathcal{U}} | \mathcal{X}_n]$, where the expectation is taken over U. Our strategy is to perform a truncation on $L_n^{\mathcal{U}}$ using events $E_n \in \mathcal{U}_n$ such that $\mathbb{P}_{0,n}[E_n^C] \to 0$ as $n \to \infty$. These truncations will enforce "well behaved" draws of U. Now, we construct a truncated proxy for our likelihood ratio, and show that the results hold for this proxy, which in turn means that they hold for the likelihood ratio test, through convergence results.

$$\tilde{L}_n := \mathbb{E}[L_n^{\mathcal{U}} \mathbb{1}_{E_n} | \mathcal{X}_n]$$

If we can provide events E_n such that we can bound the second moment of this likelihood ratio, then we will derive a condition for contiguity and in turn, we will also have the asymptotic distribution of the likelihood ratio, arising from Proposition 3. Indeed, if we construct events E_n such that

$$\operatorname*{limsup}_{n\to\infty} [\tilde{L}_n^2] = \operatorname*{limsup}_{n\to\infty} \left[\left(\frac{1}{\mathbb{P}_{0,n}[E_n]} \tilde{L}_n \right)^2 \right] \leq \exp\left\{ \sum_{\ell=1}^{\infty} \frac{\mu(\ell)^2}{v(\ell)} \right\} < \infty$$

the rest of the proof will follow by invoking Proposition 3.

Assuming such events exist, the last inequality will provide conditions on the number of spikes or the spike strength for mutual contiguity. Additionally, using the same proposition, we have

$$\frac{1}{\mathbb{P}_{0,n}[E_n]} \tilde{L}_n | \mathbb{P}_{0,n} \xrightarrow{d} \exp\left\{ \sum_{\ell=1}^{\infty} \frac{2\mu(\ell) Z_\ell - \mu(\ell)^2}{4\ell} \right\}$$

where Z_ℓ are $N(0, v(\ell))$ variables, and

$$\frac{1}{\mathbb{P}_{0,n}[E_n]} \tilde{L}_n | \mathbb{P}_{1,n} \xrightarrow{d} \exp\left\{ \sum_{\ell=1}^{\infty} \frac{2\mu(\ell) Z_\ell - \mu(\ell)^2}{4\ell} \right\}$$

where Z_ℓ are $N(\mu(\ell), v(\ell))$ variables.

We cite the following lemma, which will be proved in the Appendix, for the convergence of the likelihood ratio to the proxy.

Lemma 4 *(Convergence of proxy to likelihood ratio)*

$$L_n - \tilde{L}_n | \mathbb{P}_{0,n} \to 0$$

Thus, we can prove the necessary results for $\tilde{L}_n$, then appeal to this convergence argument to extend the results to our likelihood ratio. Now, we show that this truncated likelihood ratio follows the boundedness properties in the second moment method, using the bipartite signed cycles $B_{n,\ell}$.

Thus, we have the asymptotic distributions of $\log(L_n)$. We can see that $\sum_{\ell=1}^{\infty} \frac{\mu_\ell^2}{2\ell} > 0$, which means that the log likelihood ratio has different distributions under null and alternate. Thus, the likelihood ratio test has non-trivial power.

The only part that remains to be proven in order to invoke the proposition is that events $E_n \in \mathcal{U}_n$ exist, with the desired properties.

Now, we perform a common manipulation of second moment, using identically drawn copies of the conditional likelihood ratio.

$$\begin{aligned} \mathbb{E}_{0,n}[\tilde{L}_n^2] &= \mathbb{E}\left[\mathbb{E}[L_n^{\mathcal{U}} \mathbb{1}_{E_n} | \mathcal{X}_n]^2\right] \\ &= \mathbb{E}\left[\mathbb{E}[L_n^{\mathcal{U}(1)} L_n^{\mathcal{U}(2)} \mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}} | \mathcal{X}_n]\right] \\ &= \mathbb{E}\left[\mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}} \mathbb{E}[L_n^{\mathcal{U}(1)} L_n^{\mathcal{U}(2)} | \mathcal{U}_n]\right] \end{aligned}$$

Here, we take two independent copies of $L_n^{\mathcal{U}}$, by taking two iid copies of U (denoted as $U^{(1)}, U^{(2)}$), keeping X fixed. We can do this under the null, since W is independent of U. Using this formulation, we can calculate the following upper bound, the full derivation of which will be provided in the Appendix.

$$\begin{aligned} \mathbb{E}[L_n^{\mathcal{U}(1)} L_n^{\mathcal{U}(2)} | \mathcal{U}_n] &\leq \exp\left\{ \frac{\lambda_1^2 n}{4} \sum_{i=1}^{n} \left(\sum_{l=1}^{k_1} U_{i,l}^{(1)^2} \right) \left(\sum_{l=1}^{k_1} U_{i,l}^{(2)^2} \right) \right\} \\ &:= \psi_n(U^{(1)}, U^{(2)}) \end{aligned}$$

If we can carefully analyze $\psi_n(U^{(1)}, U^{(2)})$, then we can show the bound of the second moment of the likelihood ratio. We can use a Gaussian formulation for $U^{(1)}$ and $U^{(2)}$, that is for matrices $Z^{(1)}$ and $Z^{(2)}$ with $Z_{ij}^{(1)} \sim N(0,1)$ and $Z_{ij}^{(2)} \sim N(0,1)$, we have

$$U^{(1)} = Z^{(1)} \left(Z^{(1)^T} Z^{(1)} \right)^{-1/2} \text{ and } U^{(2)} = Z^{(2)} \left(Z^{(2)^T} Z^{(2)} \right)^{-1/2}$$

Define

$$E_{n,\alpha}^{(1)} := \left\{ \left|\left| Z^{(1)^T} Z^{(1)} - n I_n \right|\right|_{\max} \leq \frac{1}{n^{-\alpha}} \right\}$$

where I_n is the identity matrix of size n and $\alpha > 0$. We have the following lemmas related to this sequence of events.

Lemma 5 *(High probability of selected events and bounds on inverse Wishart matrix)*
Let Z be an $n \times n$ matrix of i.i.d. $N(0,1)$ entries. Define the event $E_{n,\alpha} := \left\{ \left|\left| \frac{1}{n} Z^T Z - I_n \right|\right|_{max} \leq \frac{1}{n^{1-\alpha}} \right\}$ where I_n is the identity matrix of size n and $\alpha > 0$. Then $\mathbb{P}\left((E_{n,\alpha})^C \right) \to 0$ as $n \to \infty$

Additionally, given these events $E_{n,\alpha}$, we have

$$\frac{1}{n}I_n - \frac{n^{\alpha-1}}{1+n^{1+\alpha}}J_n \preceq \left(Z^TZ\right)^{-1} \preceq \frac{1}{n}I_n + \frac{n^{\alpha-1}}{1-n^{1+\alpha}}J_n$$

where J_n is the constant matrix of size n.

We similarly define $E_n^{(2)}$. Note that for $\lambda_1 < 1$, we have that $\psi_n \mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}}$ is uniformly integrable. It is therefore sufficient to bound this quantity in order to show the regions of impossibility.

Now, under $E_n^{(1)}$ and $E_n^{(2)}$, through Lemma 5, we effectively replace $U^{(1)}$ with $\frac{1}{\sqrt{n}}Z^{(1)}$.

$$\begin{aligned}
\psi_n \mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}} &= \exp\left\{\frac{\lambda_1^2}{4}\frac{1}{n}\sum_{i=1}^n \left(\sum_{l=1}^{k_1} Z_{i,l}^{(1)^2}\right)\left(\sum_{l=1}^{k_1} Z_{i,l}^{(2)^2}\right)\right\} \\
&\to \exp\left\{\frac{\lambda_1^2}{4}\mathbb{E}\left[C^{(1)}C^{(2)}\right]\right\} \text{ where } C^{(i)} \overset{i.i.d.}{\sim} \chi^2_{k_1} \\
&= \exp\left\{\frac{k_1^2\lambda_1^2}{4}\right\} < \exp\left\{\sum_{\ell=1}^\infty \frac{\mu(\ell)^2}{v(\ell)}\right\} < \infty
\end{aligned}$$

for large enough $\lambda_1 < 1$. Thus, for $k_0 = 0$, we have shown that no test is consistent when $\lambda_1 < 1$, and the likelihood ratio test has non-trivial power for large enough values of λ_1.

The proof under the general situation where $k_1 - k_0 = \mathcal{O}(1)$ follows similarly.

Now, our testing setup is given as

$$H_0: \ X = U\Lambda_0 U^T + W \text{ vs. } H_1: \ X = U\Lambda U^T + W$$

where $\Lambda_0 = \operatorname{diag}(\lambda_0 \sum_{i=1}^{k_0} e_i), \Lambda = \operatorname{diag}(\lambda_0 \sum_{i=1}^{k_0} e_i + \lambda_1 \sum_{i=k_0+1}^{k_1} e_i)$, $U \sim \operatorname{Unif}(\mathbb{S}^n)$, $W \sim$ GOE(n), W, U are independent, $k_1 - k_0$ bounded and $\lambda_0 > \lambda_1$

This extended regime follows in a similar fashion. The likelihood ratio under this testing problem is the same as that of the case when $k_0 = 0$, with $k_1 - k_0$ spikes of size λ_1. Additionally, we construct centered lower diagonal bipartite signed cycles on the matrix X

$$\tilde{B}_{n,\ell} = 2^{\ell/2} \sum_{\substack{i_0, j_0, \ldots, i_{\ell-1}, j_{\ell-1} \\ j_a \le i_a \ \forall a \in \{1,\ldots,\ell-1\}}} X_{i_0 j_0} X_{i_1 j_0} X_{i_1 j_1} \ldots X_{i_{\ell-1} j_{\ell-1}} X_{i_0 j_{\ell-1}} - \mu(\ell)$$

where $i_0, i_1, \ldots, i_{\ell-1} \in \{1, \ldots, n\}$ all distinct and $j_0, j_1, \ldots, j_{\ell-1} \in \{1, \ldots, n\}$ all distinct and $\mu(\ell)$ is defined as the mean of the bipartite signed cycles on censored spiked GOE matrices with k_0 spikes of size λ_0. Under the alternate, the mean can be represented as $\tilde{\mu}(\ell)$ which is the mean of the bipartite signed cycles for $k_1 - k_0$ spikes of size λ_1. Therefore, for $k_1 - k_0$ bounded, if $\lambda_1 < 1$, no test is consistent, and the likelihood ratio test has non-trivial power for large enough values of λ_1.

Additionally, we have the asymptotic distribution of the likelihood ratio, given in Proposition 3. We have under the null

$$L_n^{\mathcal{U}} \xrightarrow{d} \exp\left\{\sum_{\ell=1}^\infty \frac{2\mu(\ell)Z_\ell - \mu(\ell)^2}{4\ell}\right\}$$

where Z_ℓ are $N(0, v(\ell))$ variables, and under the alternate

$$L_n^{\mathcal{U}} \xrightarrow{d} \exp\left\{\sum_{\ell=1}^{\infty} \frac{2\mu(\ell) Z_\ell - \mu(\ell)^2}{4\ell}\right\}$$

where Z_ℓ are $N(\mu(\ell), v(\ell))$ variables.

We can construct the risk of the log of the likelihood ratio test using these normal expansions. The null and alternate distributions are normal, as each element is an independent normal random variable. The variance of the log-likelihood $\tilde{v}$ is the same under the null and the alternate, while the means $-\tilde{\mu}, \tilde{\mu}$ are symmetric about 0. This implies that the optimal testing cutoff would be at the origin, and we reject the null if the log likelihood ratio is greater than 0. The risk therefore given by $2\left(1 - \Phi\left(\tilde{\mu}/\sqrt{\tilde{v}}\right)\right)$, where Φ is the Gaussian distribution function.

We have

$$\tilde{\mu} = \sum_{\ell=1}^{\infty} \frac{\mu(\ell)^2}{2\ell} = \sum_{\ell=1}^{\infty} \frac{(k_1 - k_0)\lambda_1^{2\ell} f(\ell)}{2\ell}$$

$$\tilde{v} = \sum_{\ell=1}^{\infty} \frac{\mu(\ell)^2 v(\ell)}{4\ell^2} = \sum_{\ell=1}^{\infty} \frac{(k_1 - k_0)^2 \lambda_1^{4\ell} f(\ell)^3}{2\ell}$$

We make a heuristic argument that the behaviour of these quantities is not strongly influenced by $f(\ell)$ and that gives a proper Both summations can be solved analytically using computational tools, resulting in formulation of generalized hypergeometric functions. We finally get

$$\mathrm{Risk}_{n,\lambda}(\log L_n) \leq \mathrm{erfc}\left(\frac{(k_1 - k_0)}{4}\sqrt{\log\left(\frac{1}{1-\lambda}\right)}\right)$$

4.2. Proof sketch of Theorem 2

We can show the regions of impossibility in this regime through a careful direct analysis of the likelihood ratio. We will assume throughout that $k_1 - k_0 > 2$. The proof otherwise can be recovered from the previous case considered. We are testing

$$H_0 : X = U^T \Lambda_0 U + W \text{ vs } H_1 : X = U^T \Lambda U + W$$

where W is a $\mathrm{GOE}(n)$ random matrix, $\Lambda_0 = \sum_{i=1}^{k_0} \lambda_0 e_i e_i^T$, $\Lambda = \sum_{i=1}^{k_0} \lambda_0 e_i e_i^T + \sum_{i=k_0+1}^{k_1} \lambda_1 e_i e_i^T$ and U is an $n \times n$ orthogonal matrix.

We have the following

$$f_0|U = C\exp\left\{\frac{-n}{4}\mathrm{tr}\left(\left(X - U\Lambda_0 U^T\right)^2\right)\right\}, \; f_1|U = C\exp\left\{\frac{-n}{4}\mathrm{tr}\left(\left(X - U\Lambda U^T\right)^2\right)\right\}$$

$$L_n = \frac{f_1}{f_0} = \mathop{\mathbb{E}}_{U \in \mathbb{S}^n} \exp\left\{-\frac{n}{4}\left((k_1 - k_0)\lambda_1^2 - 2\mathrm{tr}\left(XU(\Lambda - \Lambda_0)U^T\right)\right)\right\}$$

The proof will follow through the analysis of the second moment of this likelihood ratio. Finiteness of this second moment will imply mutual contiguity of the distribution under the null and alternate. We compute this second moment by taking two independent and identically distributed copies of the likelihood ratio and taking their product. We take the expectation of this second moment under the null, to get the following form:

$$\mathbb{E}\left(\left(\frac{f_1}{f_0}\right)^2\right) = \underset{U^{(a)}\in\mathbb{S}^n}{\mathbb{E}} \exp\left\{\frac{n}{2}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2 U_{ij}^{(a)^2}\right)\right\}$$

Now, this quantity is difficult to examine directly and requires spherical integrals. Thus, we use the idea of a Gaussian proxy as detailed in Ke (2016). If Z is a $\mathbb{R}^{n\times k_1}$ matrix of $N(0,1)$ variables, then we can construct $U = Z\left(Z^T Z\right)^{-1/2}$ to be uniformly drawn from the necessary Haar measure. Thus, $UU^T \stackrel{d}{=} Z\left(Z^T Z\right)^{-1} Z^T$. Also, we condition on a situation where $Z^T Z \approx nI$, akin to the proof of Theorem 1 and Lemma 5 allow us to substitute UU^T with $\frac{1}{n}ZZ^T$. We use conditioning event

$$E_{n,\alpha}^{(a)} = \left\{\left|\left|\frac{1}{n}Z^{(a)^T}Z^{(a)} - I_n\right|\right|_{\max} \leq \frac{1}{n^{1-\alpha}}\right\}$$

where I_n is the $n\times n$ identity matrix and $\alpha > 0$. We perform computations on $\mathbb{E}\left(\left(\frac{f_1}{f_0}\right)^2\right)$ by conditioning on $E_{n,\alpha}^{(a)}$ and $\left(E_{n,\alpha}^{(a)}\right)^C$, showing that the second term goes to 0 using Lemma 5, and using the same lemma to effectively replace U_{ij} with $\frac{1}{\sqrt{n}}Z_{ij}$. Following these steps gives the following bound

$$\mathbb{E}\left(\left(\frac{f_1}{f_0}\right)^2\right) \leq \underset{Z_{ij}^{(a)} \overset{i.i.d.}{\sim} \mathbb{S}^n}{\mathbb{E}} \exp\left\{\frac{1}{2}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2 Z_{ij}^{(a)^2}\right)\right\} = \left(1-\lambda_1^2\right)^{-(k_1-k_0)^2/2}$$

The final equality is from the moment generating function of a chi-squared distribution with degrees of freedom $(k_1-k_0)^2$. There is a nuance to the existence of this moment generating function, with respect to the values of λ_1 (which by definition of the regime $\lambda_1(k_1-k_0)$ being bounded and $(k_1-k_0) > 1$ exists), that we address in the full proof of this theorem in the Appendix. Now, if $\lambda_1(k_1-k_0) = \mathcal{O}(1)$, then as n goes to infinity, this term converges to

$$\exp\left\{\lim_{n\to\infty}\lambda_1^2\frac{(k_1-k_0)^2}{2}\right\} < \infty$$

if $\lambda_1\left(k_1-k_0\right) = \mathcal{O}(1)$. Thus, under this conditions, the null and alternate hypotheses are mutually contiguous. Due to the optimality of the likelihood ratio test, this shows that no test is asymptotically powerful. Additionally, we can note that if $\lambda_1\left(k_1-k_0\right) = o(1)$, then the second moment of the likelihood ratio converges in expectation to 1, which means the likelihood ratio test is asymptotically powerless, and in turn, all tests are asymptotically powerless.

5. Numerical Experiments

In this section, we consider the following simulation setups. First, we consider the subcritical regime and show that the power of the bipartite signed cycle test is close to the theoretical power limit over different secondary spike strengths and spike sparsities. Second, we compare the performance of the

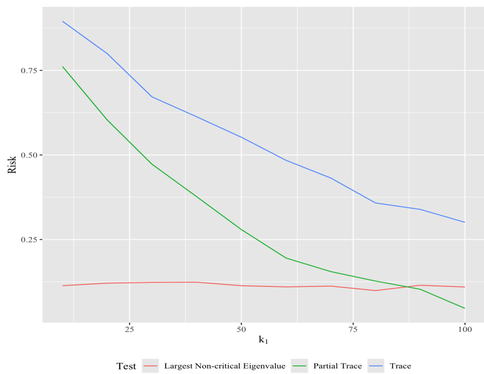

Figure 1: Risk of the bipartite signed cycle test with λ_1 for different values of k_1. Solid lines denote the theoretical limiting risk for each value of k_1

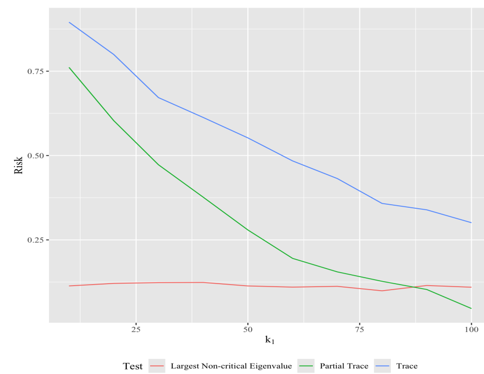

Figure 2: Comparison of the trace test ($\mathcal{T}_1$), the partial trace test ($\mathcal{T}_2$) and the largest eigenvalue test ($\mathcal{T}_3$) across different values of k_1

three tests described in Section 3 in the supercritical regime, where all the tests are asymptotically powerful.

We consider a spiked Gaussian Orthogonal Ensemble with $\lambda_0 = 2$. We vary λ_1 from 0 to 0.8. We construct 10000 samples of 50×50 GOE matrices under the null structure (to compute the empirical type-I error) and 10000 samples under the alternate structure (to compute the empirical type-II error), and we fix $k_0 = 5$. We choose different values of k_1, which give different lines in our plots. For each value of k_1, we plot the theoretical risk given by

$$\operatorname{erfc}\left(\frac{k_1 - k_0}{4}\sqrt{\log\left(\frac{1}{1-\lambda_1}\right)}\right)$$

and we plot the empirical risk of the bipartite signed cycle tests, truncating $\ell = 5$.

Figure () shows the results of this simulation, with each dotted line giving a theoretical risk and each solid line giving an empirical risk that is a good approximation of the theoretical risk.

For our comparison of tests, we consider 10000 samples of 200×200 spiked GOE matrices under the null and 10000 samples under the alternate structure, with $\lambda_0 = 2$. We fix $\lambda_1 = 1.1$ and vary k_1 from 10 to 100. For each sample, we compute the risk at each level of k_1 and plot the decay as k_1 increases. We are embodying the supercritical regime as k_1 increases, and can directly compare the small sample performance of $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{T}_3$ defined in Section 3. Figure 2 shows the performance of each estimator. The largest non-critical eigenvalue test ($\mathcal{T}_3$) has steady risk for different values of k_1. Both partial trace ($\mathcal{T}_2$) and trace test ($\mathcal{T}_1$) have decaying risk for larger values of k_1, with $\mathcal{T}_2$ having uniformly lower risk among the two tests.

6. Discussion

In this paper, we have investigated the problem of detection in a spiked Wigner matrix setup. When the number of spikes is bounded, we have sharp bounds for detection that depend solely on the strength of the secondary spikes in a GOE model. Additionally, we construct a test that attains optimal power, using bipartite signed cycles. Subsequently, using this optimal testing framework, we can perform sequential testing in order to estimate the number of spikes algorithmically. Moreover, when the number of spikes is unbounded, we also ascertain conditions for successful detection along with rate optimal tests. Extending these results to other types of spiked models considered in the seminal paper of Johnstone and Onatski (2018) remains exciting future direction. As our lower bounds are often dependent on the Gaussian Orthogonal Ensemble, a natural extension would be to generalize the results to general Wigner matrices with the necessary moments.

References

Zhidong Bai and Jian-feng Yao. Central limit theorems for eigenvalues in a spiked population model. In *Annales de l'IHP Probabilités et Statistiques*, volume 44, pages 447–474, 2008.

Zhidong Bai, Huixia Liu, and Wing-Keung Wong. Enhancement of the applicability of markowitz's portfolio optimization by utilizing random matrix theory. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 19(4):639–667, 2009.

Jinho Baik and Jack W Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6):1382–1408, 2006.

Jinho Baik, Gerard Ben Arous, and Sandrine Peche. Phase transition of the largest eigenvalue for non-null complex sample covariance matrices, 2004.

Debapratim Banerjee and Zongming Ma. Asymptotic normality and analysis of variance of log-likelihood ratios in spiked random matrix models, 2018.

Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

Florent Benaych-Georges, Alice Guionnet, and Mylène Maïda. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices, 2011.

Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, August 2001. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1013699998.

Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. The largest eigenvalues of finite rank deformation of large wigner matrices: convergence and nonuniversality of the fluctuations. 2009.

Hye Won Chung and Ji Oon Lee. Weak detection of signal in the spiked wigner model. In *International Conference on Machine Learning*, pages 1233–1241. PMLR, 2019.

Hye Won Chung, Jiho Lee, and Ji Oon Lee. Asymptotic normality of log likelihood ratio and fundamental limit of the weak detection for spiked wigner matrices. *arXiv preprint arXiv:2203.00821*, 2022.

Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, Lenka Zdeborová, et al. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Advances in Neural Information Processing Systems*, 29, 2016.

Edgar Dobriban. Sharp detection in pca under correlations: All eigenvalues matter. 2017.

Rishabh Dudeja, Songbin Liu, and Junjie Ma. Optimality of approximate message passing algorithms for spiked matrix models with rotationally invariant noise. *arXiv preprint arXiv:2405.18081*, 2024.

Ahmed El Alaoui and Michael I Jordan. Detection limits in the high-dimensional spiked rectangular model. In *Conference On Learning Theory*, pages 410–438. PMLR, 2018.

Ahmed El Alaoui, Florent Krzakala, and Michael Jordan. Fundamental limits of detection in the spiked wigner model. 2020.

Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large wigner matrices. *Communications in mathematical physics*, 272:185–228, 2007.

Svante Janson. Random regular graphs: Asymptotic distributions and contiguity. *Combinatorics, Probability and Computing*, 4(4):369–405, 1995. doi: 10.1017/S0963548300001735.

Iain M. Johnstone. High dimensional statistical inference and random matrices. (arXiv:math/0611589), November 2006. URL http://arxiv.org/abs/math/0611589. arXiv:math/0611589.

Iain M. Johnstone and Alexei Onatski. Testing in high-dimensional spiked models. (arXiv:1509.07269), February 2018. doi: 10.48550/arXiv.1509.07269. URL http://arxiv.org/abs/1509.07269. arXiv:1509.07269 [math, stat].

Ji Hyung Jung, Hye Won Chung, and Ji Oon Lee. Weak detection in the spiked wigner model with general rank. *arXiv preprint arXiv:2001.05676*, 2020.

Ji Hyung Jung, Hye Won Chung, and Ji Oon Lee. Detection of signal in the spiked rectangular models. In *International Conference on Machine Learning*, pages 5158–5167. PMLR, 2021.

Ji Hyung Jung, Hye Won Chung, and Ji Oon Lee. Detection problems in the spiked matrix models. *arXiv preprint arXiv:2301.05331*, 2023.

Zheng Ke. Detecting rare and weak spikes in large covariance matrices. 09 2016.

Antti Knowles and Jun Yin. The isotropic semicircle law and deformation of wigner matrices. *Communications on Pure and Applied Mathematics*, 66(11):1663–1749, April 2013. ISSN 1097-0312. doi: 10.1002/cpa.21450. URL http://dx.doi.org/10.1002/cpa.21450.

Antti Knowles and Jun Yin. The outliers of a deformed wigner matrix. *The Annals of Probability*, 42(5), September 2014. ISSN 0091-1798. doi: 10.1214/13-aop855. URL http://dx.doi.org/10.1214/13-AOP855.

Dmitriy Kunisky. Low coordinate degree algorithms i: Universality of computational thresholds for hypothesis testing. *arXiv preprint arXiv:2403.07862*, 2024.

Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. In *Conference on Learning Theory*, pages 1297–1301. PMLR, 2017.

Soumendu Sundar Mukherjee. Consistent model selection in the spiked wigner model via aic-type criteria. *arXiv preprint arXiv:2307.12982*, 2023.

Alexei Onatski, Marcelo J Moreira, and Marc Hallin. Asymptotic power of sphericity tests for high-dimensional data. 2013.

Alexei Onatski, Marcelo J Moreira, and Marc Hallin. Signal detection in high dimension: The multispiked case. 2014.

Aleksandr Pak, Justin Ko, and Florent Krzakala. Optimal algorithms for the inhomogeneous spiked wigner model. *Advances in Neural Information Processing Systems*, 36, 2024.

Nick Patterson, Alkes L. Price, and David Reich. Population structure and eigenanalysis. *PLoS Genetics*, 2(12):e190, 2006. ISSN 1553-7390, 1553-7404. doi: 10.1371/journal.pgen.0020190.

Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

Sandrine Péché. The largest eigenvalue of small rank perturbations of hermitian random matrices. *Probability Theory and Related Fields*, 134:127–173, 2006.

Amelia Perry, Alexander S. Wein, Afonso S. Bandeira, and Ankur Moitra. Optimality and sub-optimality of pca i: Spiked random matrix models. *The Annals of Statistics*, 46(5), October 2018. ISSN 0090-5364. doi: 10.1214/17-AOS1625. URL http://arxiv.org/abs/1807.00891. arXiv:1807.00891 [cs, math, stat].

Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8):904–909, August 2006. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng1847.

Saba Sharifi, Martin Crane, Atid Shamaie, and H Ruskin. Random matrix theory for portfolio optimization: a stability approach. *Physica A: Statistical Mechanics and its Applications*, 335 (3-4):629–643, 2004.

Emre Telatar. Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, 1999. ISSN 1541-8251. doi: 10.1002/ett.4460100604.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Eugene P Wigner. Random matrices in physics. *SIAM review*, 9(1):1–23, 1967.

Appendix A. Estimation of the Number of Spikes

Through Theorem 1, we have demonstrated an optimal test in the subcritical regime, when k_1 is bounded. This test can be equivalently constructed using bipartite signed cycles, elucidated in the proof of Theorem 1 in Section 2. Now, in order to extract the number of spiked eigenvalues, we will employ a sequential hypothesis testing framework. By setting $\lambda_0 = \lambda_1$ and $k_1 - k_0 = 1$ for all tests, we can increment k_0 until there is not enough evidence to reject the null, thus giving us a final number of spiked eigenvalues. Thus, we extract a sequence of p-values that arise from the values of the bipartite signed cycle. Using an algorithm for multiplicity considerations, such as the Benjamini-Yekutieli procedure (Benjamini and Yekutieli, 2001) for FDR control, the first hypothesis test that does not have sufficient evidence to reject the null will provide an estimate of the number of spikes. We will denote our input matrix as X. Let $\mathcal{B}_\ell(X) := \{\hat{B}_{\ell,i}\}$ refer to the set of bipartite signed cycles of length 2ℓ that lie completely in the lower triangle of X. Let $\mathcal{L} \subset \{1, \dots, n-1\}$ denote the set of possible values of ℓ, such that $|\mathcal{L}| = o(\sqrt{\log n})$. For each bipartite signed cycle length, we compute the mean and variance and using Proposition 1, we can achieve a p-value using the Cramer-Wold device to get a univariate normal distribution. By Theorem 1, we have shown that testing using these objects is equivalent to testing using the log-likelihood. Therefore, we codify this estimation procedure in Algorithm 1. This algorithm optimally extracts all the information from the hypothesis testing paradigm we have been investigating.

Data: Matrix X, Vector of cycle lengths $\mathcal{L}$
Result: Number of spikes k_1
Set initial spikes k_0 to be 0
Set initial length index i to be 1
Set multiplicity condition as FALSE
Initialize empty string of p-values P
while *multiplicity condition=FALSE* **do**
  while $i \leq |\mathcal{L}|$ **do**
    Set cycle length $\ell = \mathcal{L}[i]$
    Construct set of bipartite signed cycles $\mathcal{B}_\ell(X)$
    Compute $\hat{\mu}(\ell) = \frac{1}{|\mathcal{B}_\ell(X)|} \sum_{j=1}^{|\mathcal{B}_\ell(X)|} B_{\ell,j}$
    Compute $\hat{v}(\ell) = \frac{1}{|\mathcal{B}_\ell(X)|} \sum_{j=1}^{|\mathcal{B}_\ell(X)|} (B_{\ell,j} - \hat{\mu}(\ell))^2$
    Compute the theoretical mean for k_0 spikes $\tilde{\mu}(\ell)$
    Increment i
  end
  Compute $\hat{\mu} = \frac{1}{|\mathcal{L}|} \sum_{j=1}^{|\mathcal{L}|} (\hat{\mu}(\ell) - \tilde{\mu}(\ell))$
  Compute $\hat{v} = \frac{1}{|\mathcal{L}|^2} \sum_{j=1}^{|\mathcal{L}|} \hat{v}(\ell)$
  Compute p-value $p = 1 - \Phi\left(\hat{\mu}/\sqrt{\hat{v}}\right)$
  Add p to P
  Increment k_0
  if *P does not trigger multiplicity condition* **then**
    multiplicity condition=FALSE
  else
    multiplicity condition=TRUE
  end
end
$k_1 = k_0$

Algorithm 1: Estimating number of spikes through hypothesis testing

Remark *Algorithm 1 provides a consistent estimator for the number of spikes.*

This remark depends on the form of multiplicity consideration, but assuming consistent procedures such as Benjamini-Yekutieli, the consistency of the bipartite signed cycle test implies the consistency of the procedure.

Note that a similar procedure could be implemented using the highest non-critical eigenvalue test incrementally, which would be a more computationally efficient implementation, but would lack power in the subcritical regime. Such a procedure would also be consistent for estimating the number of spikes in the spiked Wigner setup.

Appendix B. Proof of Lemmas

B.1. Convergence of proxy to likelihood ratio

Lemma 4 states that the proxy of the likelihood ratio ($\tilde{L}_n$) converges to the likelihood ratio L_n. That is $L_n - \tilde{L}_n|\mathbb{P}_{0,n} \to 0$

We can define a probability measure

$$\tilde{\mathbb{Q}}_n(A_n) = \frac{1}{\mathbb{P}[\Omega_n]}\mathbb{E}_{\mathbb{P}_{0,n}}[\tilde{L}_n 1_{A_n}], \ \ \forall A_n \in \mathcal{X}_n$$

Using triangle inequality, and the fact that $\tilde{L}_n \le L_n$ almost surely under $\mathbb{P}_{0,n}$, we can show that the total variational distance between $\mathbb{P}_{1,n}$ and $\tilde{\mathbb{Q}}_n$ goes to 0 as $n \to \infty$. The details are given in Banerjee and Ma (2018). For any $\epsilon > 0$, let $A = \{L_n - \tilde{L}_n > \epsilon\}$. By definition of total variational distance, we have $|\mathbb{P}_{1,n}(A) - \tilde{\mathbb{Q}}_n(A)| \le d_{TV}(\mathbb{P}_{1,n}, \tilde{\mathbb{Q}}_n)$.

$$\frac{1}{\mathbb{P}_{0,n}[E_n]}\epsilon\mathbb{P}_{0,n}(A) - \frac{\mathbb{P}_{0,n}[E_n^C]}{\mathbb{P}_{0,n}[E_n]}\mathbb{P}_{1,n}(A) \le d_{TV}(\mathbb{P}_{1,n}, \tilde{\mathbb{Q}}_n)$$

Thus, we have

$$\mathbb{P}_{0,n}(A) \le \frac{1}{\epsilon}\left[d_{TV}(\mathbb{P}_{1,n}, \tilde{\mathbb{Q}}_n) + \frac{\mathbb{P}_{0,n}[E_n^C]}{\mathbb{P}_{0,n}[E_n]}\right] \to 0$$

B.2. High probability of selected events and bounds on inverse Wishart matrix

Let Z be an $n \times n$ matrix of i.i.d. $N(0,1)$ entries. Define the event $E_{n,\alpha} := \left\{\left|\left|\frac{1}{n}Z^T Z - I_n\right|\right|_{\max} \le \frac{1}{n^{1-\alpha}}\right\}$ where I_n is the identity matrix of size n and $\alpha > 0$. Then we claim that $\mathbb{P}\left((E_{n,\alpha})^C\right) \to 0$ as $n \to \infty$. Specifically, these events will hold with high probability. Indeed, each diagonal element is an independent chi-squared distribution with mean n. Moreover, each off-diagonal element is the sum of n products of two independent normal random variables. These variables, scaled by n, are all subexponential variables. Therefore, we have

$$\begin{aligned}
\mathbb{P}\left(E_{n,\alpha}\right) &= \mathbb{P}\left(\left|\left|\frac{1}{n}Z^T Z - I_n\right|\right|_{\max} \le \frac{1}{n^{1-\alpha}}\right) \\
&= \prod_{i=1}^{n}\prod_{j=1}^{i-1}\mathbb{P}\left(\left|\frac{1}{n}Z_i^T Z_j\right| \le \frac{1}{n^{1-\alpha}}\right)\prod_{i=1}^{n}\mathbb{P}\left(\left|\frac{1}{n}Z_i^T Z_i - 1\right| \le \frac{1}{n^{1-\alpha}}\right) \\
&\ge \left(1 - 2\exp\left\{-\frac{n^\alpha}{8}\right\}\right)^n \left(1 - 2\exp\left\{-\frac{n^\alpha}{4}\right\}\right)^{\frac{n(n-1)}{2}} \\
&\to 1
\end{aligned}$$

Therefore, we have shown that these are high probability events, and thus, we will leverage them in our proofs.

Now, given these events hold, we would like invoke properties related to the inverse Wishart matrix.

Based on the statement of the events $E_{n,\alpha}$, we can construct a bound for the spectral norm using the max norm. Thus, we have the following calculations on the event $E_{n,\alpha}$

$$I_n - n^\alpha J_n \preceq \frac{1}{n} Z^T Z \preceq I_n + n^\alpha J_n$$
$$\implies I_n - \frac{n^\alpha}{1+n^{1+\alpha}} J_n \preceq n \left(Z^T Z\right)^{-1} \preceq I_n + \frac{n^\alpha}{1-n^{1+\alpha}} J_n$$

Therefore, we can with high probability invoke suitable properties of the inverse Wishart matrix. Specifically, this implies that we can bound the inverse of $n(Z^T Z)^{-1}$ with an identity matrix, with a small perturbation, facilitating the proofs in this manuscript.

Appendix C. Proof of Theorem 1

We have the setup from (1) with $k_0 = 0$

$$H_0: \ X = W \text{ vs. } H_1: \ X = U \Lambda U^T + W$$

where $\Lambda = \operatorname{diag}(\lambda_1 \sum_{i=1}^{k_1} e_i)$, $U \sim \operatorname{Unif}(\mathbb{S}^n)$, $W \sim \operatorname{GOE(n)}$, W, U are independent and k_1 bounded.

The structure of the proof follows from Banerjee and Ma (2018). We construct a sequence of random variables that are "asymptotically sufficient" for the likelihood ratio, and provide some bound on the second moment of the likelihood ratio. The specifications of this bound will give us a region of impossibility for testing.

Crucially, the following proposition is invoked, a formulation of the second moment method (Proposition 1 in Banerjee and Ma (2018), which is a Gaussian form of Theorem 1 from Janson (1995)).

Proposition 3 (Banerjee and Ma (2018, Proposition 1)) *Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ be two sequences of probability measures such that for each n, both are defined on the common σ-algebra $(\Omega_n, \mathcal{F}_n)$. Suppose that for each $i \geq 1$, $W_{n,i}$ are random variables defined on $(\Omega_n, \mathcal{F}_n)$. The sequences of probability measures $\mathbb{P}_n$ and $\mathbb{Q}_n$ are mutually contiguous if the following conditions hold simultaneously:*

(i) *$\mathbb{Q}_n$ is absolutely continuous with respect to $\mathbb{P}_n$ for each n*

(ii) *For any fixed $\ell \geq 1$, one has $(W_{n,1}, \ldots, W_{n,\ell})|\mathbb{P}_n \xrightarrow{d} (Z_1, \ldots, Z_\ell)$ and $(W_{n,1}, \ldots, W_{n,\ell})|\mathbb{Q}_n \xrightarrow{d} (Z_1', \ldots, Z_\ell')$*

(iii) *$Z_i \sim N(0, \sigma_i^2)$ and $Z_i' \sim N(\mu_i, \sigma_i^2)$ are sequences of independent random variables*

(iv) *The likelihood ratio statistic $Y_n = \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}$ satisfies*

$$\limsup_{n \to \infty} \mathbb{E}_{\mathbb{P}_n}[Y_n^2] \leq \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i^2}{\sigma_i^2}\right\} < \infty$$

In addition, under these four conditions, we have that under $\mathbb{P}_n$,

$$Y_n \xrightarrow{d} \exp\left\{\sum_{i=1}^{\infty} \frac{\mu_i Z_i - \frac{1}{2}\mu_i^2}{\sigma_i^2}\right\}$$

Furthermore, given any $\epsilon, \delta > 0$ there exists a natural number $K = K(\delta, \epsilon)$ such that for any sequence n_l there is a further subsequence n_{l_m} such that

$$\limsup_{n\to\infty} \mathbb{P}_{n_{l_m}}\left(\left|\log(Y_{n_{l_m}}) - \sum_{\ell=1}^{K} \frac{2\mu_\ell W_{n_{l_m},\ell} - \mu_\ell^2}{2\sigma_\ell^2}\right| \geq \epsilon\right) \leq \delta$$

The random variables $W_{n,i}$ chosen in conjunction with Proposition 3 are known as *bipartite signed cycles*. Bipartite signed cycles of length 2ℓ are defined as

$$B_{n,\ell} = \frac{1}{n^\ell} \sum_{i_0,j_0,\ldots,i_{\ell-1},j_{\ell-1}} X_{i_0 j_0} X_{i_1 j_0} X_{i_1 j_1} \ldots X_{i_{\ell-1} j_{\ell-1}} X_{i_0 j_{\ell-1}}$$

where $i_0, i_1, \ldots, i_{\ell-1} \in \{1, \ldots, n\}$ all distinct and $j_0, j_1, \ldots, j_{\ell-1} \in \{1, \ldots, n\}$ all distinct. These random variables exploit the structure of the matrix, and are shown to abide by the stipulations of Proposition 3 based on the setup given in Banerjee and Ma (2018). However, in the testing setup we consider there are two additional structural caveats:

- Diagonal elements are of a different variance to off-diagonal elements
- The matrix is symmetric

The second point means that we cannot exploit independence between entries directly. We will make arguments on the rarity of these We circumvent this issue by constructing lower diagonal bipartite signed cycles on the matrix X In these cycles, each element is independent of each other. To account for the variance scaling, we use the following form of the bipartite signed cycles of length 2ℓ:

$$B_{n,\ell} = 2^{\ell/2} \sum_{\substack{i_0,j_0,\ldots,i_{\ell-1},j_{\ell-1} \\ j_a \leq i_a \ \forall a \in \{1,\ldots,\ell-1\}}} X_{i_0 j_0} X_{i_1 j_0} X_{i_1 j_1} \ldots X_{i_{\ell-1} j_{\ell-1}} X_{i_0 j_{\ell-1}}$$

where $i_0, i_1, \ldots, i_{\ell-1} \in \{1, \ldots, n\}$ all distinct and $j_0, j_1, \ldots, j_{\ell-1} \in \{1, \ldots, n\}$ all distinct. After performing computations of these signed cycles, we have the following proposition

Proposition 4 *Consider the testing problem (1) with $k_0 = 0$. Then for any fixed integer $l > 0$, there exists a $f(\ell)$ with $f(1) = 1$, $f(\ell)$ monotonously increases for $\ell \geq 2$ and $f(\ell) \to 1$ as $\ell \to \infty$ such that the following results hold:*

(i) Under H_0, when $1 \leq \ell_1 < \ldots < \ell_l = o(\sqrt{\log n})$,

$$\left(\frac{B_{n,\ell_1} - \frac{n+3}{\sqrt{2}}\mathbb{1}_{\ell_1=1}}{\sqrt{2\ell_1 f(\ell_1)}}, \cdots, \frac{B_{n,\ell_l}}{\sqrt{2\ell_l f(\ell_l)}}\right) \xrightarrow{d} N_l(0, I_l).$$

(ii) Under H_1, when $1 \leq \ell_1 < \ldots < \ell_l = o(\sqrt{\log n})$,

$$\left(\frac{B_{n,\ell_1} - \frac{n+3}{\sqrt{2}} \mathbb{1}_{\ell_1 = 1} - k_1 \lambda^{2\ell_1} f(\ell_1)}{\sqrt{2\ell_1 f(\ell_1)}}, \cdots, \frac{B_{n,\ell_l} - k_1 \lambda^{2\ell_l} f(\ell_l)}{\sqrt{2\ell_l f(\ell_l)}} \right) \xrightarrow{d} N_l(0, I_l).$$

We will prove this proposition in the next section, and for now, utilize this proposition to prove Theorem 1 in the case where $k_0 = 0$.

Using the density of a Gaussian Orthogonal Ensemble, we can construct the likelihood ratio as a function of the observed data, and the underlying orthogonal matrix. Formally, we define σ-fields $\mathcal{X}_n = \sigma(\{X_{i*}\}_{i=1}^n)$ and $\mathcal{U}_n = \sigma(\{U_{i*}\}_{i=1}^n)$ and $M_{i,j} = \sum_{l=1}^{k_1} \lambda_1 U_{i,l} U_{j,l}$. The probability measures for the null and alternate hypotheses in our analysis are given by $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ respectively. We define a conditional likelihood ratio as

$$L_n^{\mathcal{U}} := \exp\left\{ \frac{n}{2} \sum_{i=1}^n \sum_{j=1}^n \left(X_{i,j} M_{i,j} - \frac{1}{2} M_{i,j}^2 \right) \right\}$$

We can see that our likelihood ratio is given as $L_n = \mathbb{E}[L_n^{\mathcal{U}} | \mathcal{X}_n]$, where the expectation is taken over U. Our strategy is to perform a truncation on $L_n^{\mathcal{U}}$ using events $E_n \in \mathcal{U}_n$ such that $\mathbb{P}_{0,n}[E_n^C] \to 0$ as $n \to \infty$. These truncations will enforce "well behaved" draws of U. Now, we construct a truncated proxy for our likelihood ratio.

$$\tilde{L}_n := \mathbb{E}[L_n^{\mathcal{U}} \mathbb{1}_{E_n} | \mathcal{X}_n]$$

If we can prove the necessary results for $\tilde{L}_n$, then we can use some convergence arguments to extend the results to our likelihood ratio. Now, we show that this truncated likelihood ratio follows the boundedness properties in the second moment method, using the bipartite signed cycles $B_{n,\ell}$.

Now, using Proposition 4, we have that for any fixed $l \in \mathbb{N}$ and any $1 \leq \ell_1 < \ldots < \ell_l = o(\sqrt{\log n})$, under $\tilde{\mathbb{Q}}_n$,

$$\left(\frac{B_{n,\ell_1} - \frac{n+3}{\sqrt{2}} \mathbb{1}_{\ell_1 = 1} - k_1 \lambda^{2\ell_1} f(\ell_1)}{\sqrt{2\ell_1 f(\ell_1)}}, \cdots, \frac{B_{n,\ell_l} - k_1 \lambda^{2\ell_l} f(\ell_l)}{\sqrt{2\ell_l f(\ell_l)}} \right) \xrightarrow{d} N_l(0, I_l)$$

Now, let us assume that there are E_n such that

$$\limsup_{n\to\infty} [\tilde{L}_n^2] = \limsup_{n\to\infty} \left[\left(\frac{1}{\mathbb{P}_{0,n}[E_n]} \tilde{L}_n \right)^2 \right] \leq \exp\left\{ \sum_{\ell=1}^{\infty} \frac{\mu(\ell)^2}{v(\ell)} \right\} < \infty$$

The last inequality will impose conditions on λ_1. By the proposition for the second moment method, we have mutual contiguity between the null and alternate hypotheses, which means that no test can be consistent. Additionally, using the same proposition, we have

$$\frac{1}{\mathbb{P}_{0,n}[E_n]} \tilde{L}_n | \mathbb{P}_{0,n} \xrightarrow{d} \exp\left\{ \sum_{\ell=1}^{\infty} \frac{2\mu(\ell) Z_\ell - \mu(\ell)^2}{4\ell} \right\}$$

where Z_ℓ are $N(0, v(\ell))$ variables, and

$$\frac{1}{\mathbb{P}_{0,n}[E_n]} \tilde{L}_n | \mathbb{P}_{1,n} \xrightarrow{d} \exp\left\{ \sum_{\ell=1}^{\infty} \frac{2\mu(\ell) Z_\ell - \mu(\ell)^2}{4\ell} \right\}$$

where Z_ℓ are $N(\mu(\ell), v(\ell))$ variables. By Slutsky's Theorem, the same convergence works $\tilde{L}_n$.

From Lemma 4, we have that $L_n - \tilde{L}_n | \mathbb{P}_{0,n} \to 0$.

Thus, we have the asymptotic distributions of $\log(L_n)$. We can see that $\sum_{\ell=1}^{\infty} \frac{\mu_\ell^2}{2\ell} > 0$, which means that the log likelihood ratio has different distributions under null and alternate. Thus, the likelihood ratio test has non-trivial power.

The only part that remains to be proven is that events $E_n \in \mathcal{U}_n$ exist, with the desired properties. Now,

$$\begin{aligned}
\mathbb{E}_{0,n}[\tilde{L}_n^2] &= \mathbb{E}\left[\mathbb{E}[L_n^{\mathcal{U}} \mathbb{1}_{E_n} | \mathcal{X}_n]^2\right] \\
&= \mathbb{E}\left[\mathbb{E}[L_n^{\mathcal{U}(1)} L_n^{\mathcal{U}(2)} \mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}} | \mathcal{X}_n]\right] \\
&= \mathbb{E}\left[\mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}} \mathbb{E}[L_n^{\mathcal{U}(1)} L_n^{\mathcal{U}(2)} | \mathcal{U}_n]\right]
\end{aligned}$$

Here, we take two independent copies of $L_n^{\mathcal{U}}$, by taking two iid copies of U (denoted as $U^{(1)}, U^{(2)}$, keeping X fixed. We can do this under the null, since W is independent of U. We extend the σ-field $\mathcal{U}_n$ to be the field generated by both of these copies.

$$\begin{aligned}
\mathbb{E}[L_n^{\mathcal{U}(1)} L_n^{\mathcal{U}(2)} | \mathcal{U}_n] &= \exp\left\{\frac{n}{4}\sum_{i=1}^{n} M_{ii}^{(1)} M_{ii}^{(2)} - \frac{n}{8}\sum_{i \neq j}^{n}\sum^{n}\left(M_{ij}^{(1)} - M_{ij}^{(2)}\right)^2\right\} \\
&= \exp\left\{\frac{\lambda_1^2 n}{4}\sum_{i=1}^{n}\left(\sum_{l=1}^{k_1} U_{i,l}^{(1)^2}\right)\left(\sum_{l=1}^{k_1} U_{i,l}^{(2)^2}\right)\right. \\
&- \left.\frac{\lambda_1^2 n}{8}\sum_{i \neq j}^{n}\sum^{n}\left[\left(\sum_{l=1}^{k_1} U_{i,l}^{(1)} U_{j,l}^{(1)}\right) - \left(\sum_{l=1}^{k_1} U_{i,l}^{(2)} U_{j,l}^{(2)}\right)\right]^2\right\} \\
&\leq \exp\left\{\frac{\lambda_1^2 n}{4}\sum_{i=1}^{n}\left(\sum_{l=1}^{k_1} U_{i,l}^{(1)^2}\right)\left(\sum_{l=1}^{k_1} U_{i,l}^{(2)^2}\right)\right\} \\
&:= \psi_n(U^{(1)}, U^{(2)})
\end{aligned}$$

If we can carefully analyze $\psi_n(U^{(1)}, U^{(2)})$, then we can show the bound of the second moment of the likelihood ratio. We can use a Gaussian formulation for $U^{(1)}$ and $U^{(2)}$, that is for matrices $Z^{(1)}$ and $Z^{(2)}$ with $Z_{ij}^{(1)} \sim N(0,1)$ and $Z_{ij}^{(2)} \sim N(0,1)$, we have

$$U^{(1)} = Z^{(1)}\left(Z^{(1)^T} Z^{(1)}\right)^{-1/2} \text{ and } U^{(2)} = Z^{(2)}\left(Z^{(2)^T} Z^{(2)}\right)^{-1/2}$$

Define

$$E_{n,\alpha}^{(1)} := \left\{\left|\left|\frac{1}{n} Z^{(1)^T} Z^{(1)} - I_n\right|\right|_{\max} \leq \frac{1}{n^{1-\alpha}}\right\}$$

where I_n is the identity matrix of size n and $\mathbb{P}_{0,n}((E_n^{(1)})^C) \to 0$ as $n \to \infty$. This follows from the subexponential tails of the chi-squared distribution, and therefore, we can concentrate each diagonal element and off-diagonal element. We similarly define $E_n^{(2)}$. Note that for $\lambda_1 < 1$, we have that

$\psi_n \mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}}$ is uniformly integrable. Thus, bounding this quantity is sufficient to complete this proof.

Now, under $E_n^{(1)}$ and $E_n^{(2)}$, we effectively replace $U^{(1)}$ with $\frac{1}{\sqrt{n}} Z^{(1)}$.

$$
\begin{aligned}
\psi_n \mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}} &= \exp\left\{ \frac{\lambda_1^2 n}{4} \sum_{i=1}^{n} \left(\sum_{l=1}^{k_1} U_{i,l}^{(1)^2} \right) \left(\sum_{l=1}^{k_1} U_{i,l}^{(2)^2} \right) \right\} \mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}} \\
&= \exp\left\{ \frac{\lambda_1^2}{4n} \sum_{i=1}^{n} \left(\sum_{l=1}^{k_1} \left(Z_{i,l}^{(1)} \left(Z_{i,l}^{(1)^T} Z_{i,l}^{(1)} \right)^{-1} \right)^2 \right) \left(\sum_{l=1}^{k_1} \left(Z_{i,l}^{(2)} \left(Z_{i,l}^{(2)^T} Z_{i,l}^{(2)} \right)^{-1} \right)^2 \right) \right\} \mathbb{1}_{E_n^{(1)}} \mathbb{1}_{E_n^{(2)}} \\
&\leq \exp\left\{ \frac{\lambda_1^2}{4n} \sum_{i=1}^{n} \left(\sum_{l=1}^{k_1} Z_{i,l}^{(1)^2} \right) \left(\sum_{l=1}^{k_1} Z_{i,l}^{(2)^2} \right) \right\} \text{ by Lemma 5} \\
&= \exp\left\{ \frac{\lambda_1^2}{4} \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{l=1}^{k_1} Z_{i,l}^{(1)^2} \right) \left(\sum_{l=1}^{k_1} Z_{i,l}^{(2)^2} \right) \right\} \\
&\to \exp\left\{ \frac{\lambda_1^2}{4} \mathbb{E}\left[C^{(1)} C^{(2)} \right] \right\} \text{ where } C^{(i)} \overset{i.i.d.}{\sim} \chi^2_{k_1} \\
&= \exp\left\{ \frac{k_1^2 \lambda_1^2}{4} \right\} < \exp\left\{ \sum_{\ell=1}^{\infty} \frac{\mu(\ell)^2}{v(\ell)} \right\} < \infty
\end{aligned}
$$

for large enough $\lambda_1 < 1$. Thus, for $k_0 = 0$, we have shown that no test is consistent when $\lambda_1 < 1$, and the likelihood ratio test has non-trivial power for large enough values of λ_1.

For the general situation where $k_1 - k_0 = \mathcal{O}(1)$, we follow a similar procedure.

Now, our testing setup is given as

$$
H_0 : \; X = U\Lambda_0 U^T + W \text{ vs. } H_1 : \; X = U\Lambda U^T + W
$$

where $\Lambda_0 = \text{diag}(\lambda_0 \sum_{i=1}^{k_0} e_i), \Lambda = \text{diag}(\lambda_0 \sum_{i=1}^{k_0} e_i + \lambda_1 \sum_{i=k_0+1}^{k_1} e_i)$, $U \sim \text{Unif}(\mathbb{S}^n)$, $W \sim \text{GOE(n)}$, W, U are independent, $k_1 - k_0$ bounded and $\lambda_0 > \lambda_1$

The likelihood ratio under this testing problem is the same as that of the case when $k_0 = 0$, with $k_1 - k_0$ spikes of size λ_1. As we are working with isotropic matrices, we can choose a rotation such that the first k_0 eigenvectors are the canonical basis vectors. We modify our statistics to be centered lower diagonal bipartite signed cycles on the matrix X

$$
\tilde{B}_{n,\ell} = \sum_{\substack{i_0, j_0, \ldots, i_{\ell-1}, j_{\ell-1} \\ j_a \leq i_a \; \forall a \in \{1,\ldots,\ell-1\}}} X_{i_0 j_0} X_{i_1 j_0} X_{i_1 j_1} \ldots X_{i_{\ell-1} j_{\ell-1}} X_{i_0 j_{\ell-1}} - \mu(\ell)
$$

where $i_0, i_1, \ldots, i_{\ell-1} \in \{1, \ldots, n\}$ all distinct and $j_0, j_1, \ldots, j_{\ell-1} \in \{1, \ldots, n\}$ all distinct and $\mu(\ell)$ is defined as the mean of the bipartite signed cycles for k_0 spikes of size λ_0. Under the alternate, the mean can be represented as $\tilde{\mu}(\ell)$ which is the mean of the bipartite signed cycles for $k_1 - k_0$ spikes of size λ_1. Therefore, for $k_1 - k_0$ bounded, if $\lambda_1 < 1$, no test is consistent, and the likelihood ratio test has non-trivial power for large enough values of λ_1.

Additionally, we have the asymptotic distribution of the likelihood ratio, given in Proposition 3. We have under the null

$$
L_n^{\mathcal{U}} \xrightarrow{d} \exp\left\{ \sum_{\ell=1}^{\infty} \frac{2\mu(\ell) Z_\ell - \mu(\ell)^2}{4\ell} \right\}
$$

where Z_ℓ are $N(0, v(\ell))$ variables, and under the alternate

$$L_n^{\mathcal{U}} \xrightarrow{d} \exp\left\{\sum_{\ell=1}^{\infty} \frac{2\mu(\ell)Z_\ell - \mu(\ell)^2}{4\ell}\right\}$$

where Z_ℓ are $N(\mu(\ell), v(\ell))$ variables.

We can construct the risk of the log of the likelihood ratio test using these normal expansions. The null and alternate distributions are normal, as each element is an independent normal random variable. The variance of the log-likelihood $\tilde{v}$ is the same under the null and the alternate, while the means $-\tilde{\mu}, \tilde{\mu}$ are symmetric about 0. Here we have

$$\tilde{\mu} = \sum_{\ell=1}^{\infty} \frac{\mu(\ell)^2}{2\ell} = \sum_{\ell=1}^{\infty} \frac{(k_1 - k_0)\lambda_1^{2\ell} f(\ell)}{2\ell}$$

$$\tilde{v} = \sum_{\ell=1}^{\infty} \frac{\mu(\ell)^2 v(\ell)}{4\ell^2} = \sum_{\ell=1}^{\infty} \frac{(k_1 - k_0)^2 \lambda_1^{4\ell} f(\ell)^3}{2\ell}$$

This implies that the optimal testing cutoff would be at the origin, and we reject the null if the log likelihood ratio is greater than 0. The risk therefore given by $2\left(1 - \Phi\left(\tilde{\mu}/\sqrt{\tilde{v}}\right)\right)$, where Φ is the Gaussian distribution function.

Appendix D. Proof of Proposition 4

We perform an analysis of the bipartite signed cycles. In this section we use k to denote the number of secondary spikes, which is usually denoted by $k_1 - k_0$.

First, we investigate the distribution of $B_{n,1}$ under the null. Note that

$$B_{n,1} = \sqrt{2}\sum_{i \geq j}^{n}\sum^{n} X_{ij}^2 = \sqrt{2}\sum_{i=1}^{n} X_{ii}^2 + \sqrt{2}\sum_{i>j}^{n}\sum^{n} X_{ij}^2$$

We see that the squares of the matrix values are scaled chi-squared variables.

$$\text{For } i \neq j,\ \mathbb{E}[X_{ij}^2] = \frac{1}{n},\ \operatorname{Var}(X_{ij}^2) = \frac{2}{n^2}$$

$$\mathbb{E}[X_{ii}^2] = \frac{2}{n},\ \operatorname{Var}(X_{ii}^2) = \frac{8}{n^2}$$

$$\implies \mathbb{E}[B_{n,1}] = 2\sqrt{2} + \frac{n-1}{\sqrt{2}} = \frac{n+3}{\sqrt{2}},\ \operatorname{Var}(B_{n,1}) = \frac{16}{n} + \frac{2(n-1)}{n} = \frac{n+7}{n} \to 2$$

Now, let us look at the distribution of $B_{n,\ell}$ where $\ell \geq 2$. We see that the expectation is 0 since each of the elements are mean 0 and independent of each other. Let us denote the set of lower diagonal bipartite signed cycles of length 2ℓ as $\mathcal{B}_{n,\ell}$, with each cycle being denoted by ω. For a given cycle ω, we index the product of its elements as X_ω. Additionally, we denote set of lower diagonal bipartite signed cycles including d diagonal elements in the cycle, as $\mathcal{B}_{n,\ell}^d$, and we see that $\mathcal{B}_{n,\ell} = \mathcal{B}_{n,\ell}^0 \cup \mathcal{B}_{n,\ell}^1 \cup \ldots \cup \mathcal{B}_{n,\ell}^{\ell-1}$, with each set disjoint.

$$
\begin{aligned}
\mathrm{Var}(B_{n,\ell}) &= \mathbb{E}[B_{n,\ell}^2] \\
&= \mathbb{E}\left[\left(\sum_{\omega\in\mathcal{B}_{n,\ell}} X_\omega\right)^2\right] \\
&= \mathbb{E}\left[\sum_{\omega,\omega^{'}\in\mathcal{B}_{n,\ell}} X_\omega X_{\omega^{'}}\right]
\end{aligned}
$$

The value inside the summand is non-zero only if the two cycles ω and $\omega^{'}$ have all the same elements. For a given ω, there are 2ℓ possibilities for $\omega^{'}$. The variance itself will depend on the number of diagonal elements in the cycle ω. Thus, we have

$$
\mathrm{Var}(B_{n,\ell}) = \frac{2^{\ell+1}\ell}{n^{2\ell}} \sum_{d=0}^{\ell-1} \left|\mathcal{B}_{n,\ell}^d\right| 2^d
$$

Additionally, the notion of asymptotic normality follows from the application of Wick's theorem in Banerjee and Ma (2018). A similar argument can be made in the case of these modified bipartite signed cycles.

Under the alternate, let us denote $M_{ij} = e_i^T U\Lambda U^T e_j = \sum_{l=1}^{X} \lambda U_{il}U_{jl}$. Note that $\sum_{i=1}^n \sum_{j=1}^n M_{ij}^2 = \mathrm{tr}\left(\left(U\Lambda U^T\right)^2\right) = X\lambda^2$

$$
\mathbb{E}\left[X_{ij}^2|U\right] = \frac{1}{n} + M_{ij}^2,\ \mathbb{E}\left[X_{ii}^2|U\right] = \frac{2}{n} + M_{ii}^2
$$

$$
\mathrm{Var}\left(X_{ij}^2|U\right) = \frac{2}{n^2} + \frac{4M_{ij}^2}{n},\ \mathrm{Var}\left(X_{ii}^2|U\right) = \frac{8}{n^2} + \frac{8M_{ij}^2}{n}
$$

$$
\mathbb{E}\left[B_{n,1}|U\right] = \frac{n+3}{\sqrt{2}} + \frac{k\lambda^2}{\sqrt{2}} + \frac{1}{\sqrt{2}}\sum_{i=1}^n M_{ii}^2
$$

$$
\mathrm{Var}\left(B_{n,1}\right) = \frac{2(n+7)}{n} + \frac{4k\lambda^2}{n} + \frac{12}{n}\sum_{i=1}^n M_{ii}^2
$$

Using a Gaussian approximation, we have $\sum_{i=1}^n M_{ii}^2 \approx \frac{2k\lambda^2}{n}$. Thus, asymptotically, we have

$$
\mathbb{E}\left[B_{n,1}\right] \approx \frac{n+3}{\sqrt{2}} + \frac{k\lambda^2}{\sqrt{2}}
$$

$$
\mathrm{Var}\left(B_{n,1}\right) = \frac{2(n+7)}{n} \to 2
$$

The analysis of $B_{n,\ell}$ follows similar to the analysis in Banerjee and Ma (2018). We split the summand $B_{n,\ell}$ as follows

$$
\begin{aligned}
B_{n,\ell} &= 2^{\ell/2}\sum X_{i_0j_0}X_{i_0j_1}\dots X_{i_{k-1}j_0} \\
&= 2^{\ell/2}\sum W_{i_0j_0}W_{i_0j_1}\dots W_{i_{k-1}j_0} + \mu(\ell)d + V_{n,\ell}
\end{aligned}
$$

where μ_ℓ is the summation with the elements M_{ij} and $V_{n,\ell}$ contains all the cross terms with the scaling. The first summation has the same properties $B_{n,\ell}$ under the null. Also, the variance of $B_{n,\ell}$ is determined by just the first term, and thus, is the same as the null case. The analysis of $V_{n,\ell}$ follows similarly to Banerjee and Ma (2018), and it is shown to be a negligible term in the mean and variance computations.

We look at the expectation of $B_{n,\ell}$ conditioned on U.

$$\mathbb{E}\left[B_{n,\ell}|U\right] \;=\; \mathbb{E}\left[\sum_{\omega\in\mathcal{B}_{n,\ell}} M_\omega|U\right] =: \mu(\ell)$$

Now, we have $M_{ij} = \sum_{l=1}^{X} U_{il}U_{jl}$. The product of these matrix values have a non-zero expectation iff every element has at least two copies of U_{ij}. We see that for a fixed l, each row and column is represented twice in the cycle, and using a Gaussian approximation, will have an expectation of $\frac{1}{n^{2\ell}}$. This expectation doubles for every overlap between $\{i_0, i_1, \ldots, i_{\ell-1}\}$ and $\{j_0, j_1, \ldots, j_{\ell-1}\}$, as each overlap counts the given element 4 times. The distribution of the number of overlaps is a hypergeometric distribution with parameters (n, ℓ, ℓ). Thus, we have

$$\mu(\ell) = \frac{k2^\ell\lambda^{2\ell}}{n^{2\ell}}|\mathcal{B}_{n,\ell}|G_\ell(2)$$

where $G_\ell(z)$ is the probability generating function of a hypergeometric distribution with parameters (n, ℓ, ℓ) evaluated at z. The analysis of $|\mathcal{B}_{n,\ell}|$ is given in the later part of this section.

We also provide an analysis of $|\mathcal{B}^d_{n,\ell}|$. We provide a probabilistic heuristic for this calculation. For a cycle ω in $\mathcal{B}_{n,\ell}$, the number of diagonal elements depends on the number of neighboring elements sharing a main diagonal. Therefore, if we think about two points chosen to be two away in the cycle, we denote the probability that they belong to the same diagonal as p. Therefore, we have an approximate size of $|\mathcal{B}^d_{n,\ell}| = |\mathcal{B}_{n,\ell}|\binom{\ell-1}{d}p^d(1-p)^{\ell-1-d}$, as we treat the number of diagonal elements as a binomial variable. Therefore, our variance estimate is of the form

$$\mathrm{Var}(B_{n,\ell}) = \frac{2^{\ell+1}\ell}{n^{2\ell}}\,|\mathcal{B}_{n,\ell}|\sum_{d=0}^{\ell-1}\binom{\ell-1}{d}p^d(1-p)^{\ell-1-d}2^d = \frac{2^{\ell+1}\ell}{n^{2\ell}}\,|\mathcal{B}_{n,\ell}|\,(1+p)^{\ell-1}$$

The probability p can be worked out to be $\frac{2(2n-1)}{n(n+1)}$. We also note that $\frac{G(\ell)^2}{(1+p)^{\ell-1}} > 1$ and converges to 1 as n goes to infinity. We make a DCT-type argument to continue the calculations without this term.

We derive a recurrence relation for $|\mathcal{B}_{n,\ell}|$ and reason about its asymptotic growth in Appendix **??**. We show that

$$|\mathcal{B}_{n,\ell}| = (1+o(1))\frac{n^{2\ell}}{2^\ell}f(\ell)$$

where $f(\ell) \uparrow 1$ as $\ell \to \infty$.

Thus, we have Thus, we have

$$\frac{\mu(\ell)^2}{v(\ell)} = \frac{k^2\lambda^{4l}f(\ell)(G(\ell)^2)}{2\ell(1+p)^{\ell+1}}$$

We can see by the ratio test that if $\lambda < 1$, then $\sum_{\ell=1}^{\infty}\frac{\mu(\ell)^2}{v(\ell)} < \infty$. Combining these results, we have the proposition regarding bipartite signed cycles.

Appendix E. Proof of Theorem 2

We prove this theorem through the direct analysis of the likelihood ratio and use techniques similar to the ones described in Ke (2016) for the regime where k_1 is unbounded but allows for unbounded k_0 as well whereas Ke (2016) considers only the case of $k_0 = 0$ and polynomial scaling in dimension of k_1. Similar analysis of the second moment of likelihoods are given in Perry et al. (2018). We are testing

$$H_0 : X = U^T \Lambda_0 U + W \text{ vs } H_1 : X = U^T \Lambda U + W$$

where W is a (n) random matrix, $\Lambda_0 = \sum_{i=1}^{k_0} \lambda_0 e_i e_i^T$, $\Lambda = \sum_{i=1}^{k_0} \lambda_0 e_i e_i^T + \sum_{i=k_0+1}^{k_1} \lambda_1 e_i e_i^T$ and U is an $n \times n$ orthogonal matrix.

We have the following

$$\begin{aligned}
f_0|U &= C \exp\left\{\frac{-n}{4}\mathrm{tr}\left(\left(X - U\Lambda_0 U^T\right)^2\right)\right\} \\
f_1|U &= C \exp\left\{\frac{-n}{4}\mathrm{tr}\left(\left(X - U\Lambda U^T\right)^2\right)\right\} \\
L_n &= \frac{f_1}{f_0} \\
&= \mathop{\mathbb{E}}_{U \in \mathbb{S}^n} \exp\left\{-\frac{n}{4}\left((k_1 - k_0)\lambda_1^2 - 2\mathrm{tr}\left(XU(\Lambda - \Lambda_0)U^T\right)\right)\right\}
\end{aligned}$$

We will show the conditions for contiguity appealing to the Second Moment Method, that is used to show lower bounds in composite hypothesis testing problems. If we can show that the second moment of the likelihood ratio is finite, that implies a concentration of the likelihood ratio as n goes to infinity. This in turn implies that the power of the likelihood ratio test is strictly bounded away from 1, implying the likelihood ratio test is not asymptotically powerful. Equivalently, we can say that the distributions under the null and alternate hypotheses are mutually contiguous. In order to construct the second moment of the likelihood ratio, we take two independent and identically distributed copies of the likelihood ratio and multiply them.

We take expectation over $X - U\Lambda_0 U^T$ under the null, to get

$$
\begin{aligned}
&\mathbb{E}\left(\left(\frac{f_1}{f_0}\right)^2\right) = \\
&\mathop{\mathbb{E}}_{U^{(1)},U^{(2)}\in\mathbb{S}^n} \mathbb{E}_{H_0} \exp\left\{\frac{n}{2}\mathrm{tr}\left(X\left(U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T} + U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}\right)\right)\right.\\
&\quad\left. - \frac{n}{4}\left(\mathrm{tr}\left(U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T}U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T} + U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}\right)\right)\right\}\\
&= \mathop{\mathbb{E}}_{U^{(1)},U^{(2)},U^{(3)}\in\mathbb{S}^n} \mathbb{E}_{H_0} \exp\left\{\frac{n}{2}\mathrm{tr}\left(\left(X - U^{(3)}\Lambda_0 U^{(3)^T}\right)\left(U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T} + U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}\right)\right)\right.\\
&\quad + \frac{n}{2}\left(\mathrm{tr}\left(U^{(3)}\Lambda_0U^{(3)^T}U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T} + U^{(3)}\Lambda_0U^{(3)^T}U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}\right)\right)\\
&\quad\left. - \frac{n}{4}\left(\mathrm{tr}\left(U^{(1)}(\Lambda-\Lambda_0)^2U^{(1)^T} + U^{(2)}(\Lambda-\Lambda_0)^2U^{(2)^T}\right)\right)\right\}\\
&= \mathop{\mathbb{E}}_{U^{(1)},U^{(2)},U^{(3)}\in\mathbb{S}^n} \exp\left\{\frac{n}{2}\left(\mathrm{tr}\left(U^{(3)}\Lambda_0U^{(3)^T}U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T} + U^{(3)}\Lambda_0U^{(3)^T}U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}\right)\right)\right.\\
&\quad\left. - \frac{n}{4}\left(\mathrm{tr}\left(U^{(1)}(\Lambda-\Lambda_0)^2U^{(1)^T} + U^{(2)}(\Lambda-\Lambda_0)^2U^{(2)^T}\right)\right)\right\}\\
&\quad \times \exp\left\{\frac{n}{4}\mathrm{tr}\left(\left(U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T} + U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}\right)\left(U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T} + U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}\right)\right)\right\}\\
&= \mathop{\mathbb{E}}_{U^{(1)},U^{(2)},U^{(3)}\in\mathbb{S}^n} \exp\left\{\frac{n}{2}\left(\mathrm{tr}\left(U^{(3)}\Lambda_0U^{(3)^T}U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T} + U^{(3)}\Lambda_0U^{(3)^T}U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}\right.\right.\right.\\
&\quad\left.\left.\left. + U^{(1)}(\Lambda-\Lambda_0)U^{(1)^T}U^{(2)}(\Lambda-\Lambda_0)U^{(2)^T}\right)\right)\right\}\\
&= \mathop{\mathbb{E}}_{U^{(a)}\in\mathbb{S}^n} \exp\left\{\frac{n}{2}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2 U_{ij}^{(a)^2}\right)\right\}
\end{aligned}
$$

The first two terms are 0 as there is no overlap between the non-zero elements. Now, this quantity is difficult to examine directly and requires spherical integrals. Thus, we use the idea of a Gaussian proxy as detailed in Ke (2016). If Z is a $\mathbb{R}^{n\times k_1}$ matrix of $N(0,1)$ variables, then we can construct $U = Z\left(Z^TZ\right)^{-1/2}$ to be uniformly drawn from the necessary Haar measure. Thus, $UU^T \stackrel{d}{=} Z\left(Z^TZ\right)^{-1}Z^T$. Also, we condition on a situation where $Z^TZ \approx nI$, akin to the proof of Theorem 1, thus we use $UU^T \approx \frac{1}{n}ZZ^T$. We use conditioning event

$$
E_{n,\alpha}^{(a)} = \left\{\left|\left|\frac{1}{n}Z^{(a)^T}Z^{(a)} - I_n\right|\right|_{\max} \le \frac{1}{n^{1-\alpha}}\right\}
$$

$$
\begin{aligned}
\mathbb{E}\left(\left(\frac{f_1}{f_0}\right)^2\right) &= \mathbb{E}\left(\left(\frac{f_1}{f_0}\right)^2\right)\mathbb{1}_{E_{n,\alpha}^{(a)}} + \mathbb{E}\left(\left(\frac{f_1}{f_0}\right)^2\right)\mathbb{1}_{\left(E_{n,\alpha}^{(a)}\right)^C} \\
\mathbb{E}\left(\left(\frac{f_1}{f_0}\right)^2\right)\mathbb{1}_{E_{n,\alpha}^{(a)}} &= \mathop{\mathbb{E}}_{U^{(a)}\in\mathbb{S}^n} \exp\left\{\frac{n}{2}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2 U_{ij}^{(a)^2}\right)\right\}\mathbb{1}_{E_{n,\alpha}^{(a)}} \\
&= \mathop{\mathbb{E}}_{Z_{ij}^{(a)}\overset{i.i.d.}{\sim}\mathbb{S}^n} \exp\left\{\frac{n}{2}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2\left(Z_{ij}^{(a)}\left(Z_{ij}^{(a)^T}Z_{ij}^{(a)}\right)^{-1}\right)^2\right)\right\}\mathbb{1}_{E_{n,\alpha}^{(a)}} \\
&\leq \mathop{\mathbb{E}}_{Z_{ij}^{(a)}\overset{i.i.d.}{\sim}\mathbb{S}^n} \exp\left\{\frac{1}{2}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2 Z_{ij}^{(a)^2}\right)\right\} \text{ by Lemma 5} \\
&= \left(1-\lambda_1^2\right)^{-(k_1-k_0)^2/2}
\end{aligned}
$$

The final equality arises from the moment generating function of a chi-squared random variable with $(k_1 - k_0)^2$ degrees of freedom. For this moment generating function, we require $\lambda_1 < 1$. As we are operating under the assumption that $\lambda_1(k_1 - k_0) < \infty$, for large enough n, we have $\lambda_1 < 1$. Therefore, we can asymptotically reason that this form of the second moment is well defined. We must show that the second term goes to 0 using Lemma 5.

$$
\begin{aligned}
\mathbb{E}\left(\left(\frac{f_1}{f_0}\right)^2\right)\mathbb{1}_{\left(E_{n,\alpha}^{(a)}\right)^C} &= \mathop{\mathbb{E}}_{U^{(a)}\in\mathbb{S}^n} \exp\left\{\frac{n}{2}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2 U_{ij}^{(a)^2}\right)\right\}\mathbb{1}_{\left(E_{n,\alpha}^{(a)}\right)^C} \\
&= \mathop{\mathbb{E}}_{Z_{ij}^{(a)}\overset{i.i.d.}{\sim}\mathbb{S}^n} \exp\left\{\frac{n}{2}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2\left(Z_{ij}^{(a)}\left(Z_{ij}^{(a)^T}Z_{ij}^{(a)}\right)^{-1}\right)^2\right)\right\}\mathbb{1}_{\left(E_{n,\alpha}^{(a)}\right)^C}
\end{aligned}
$$

Now, under $\left(E_{n,\alpha}^{(a)}\right)^C$, we know that there exists i, j such that $Z_i^{(a)T}Z_j^{(a)} \geq n\mathbb{1}_{i=j} + n^{\alpha}$. Therefore, we have the largest eigenvalue of $Z^{(a)T}Z^{(a)}$ is bounded below by $n^{1+\alpha}$. Therefore, the smallest eigenvalue of $\left(Z^{(a)T}Z^{(a)}\right)^{-1}$ is bounded below by $\frac{1}{n^{1-\alpha}}$. Therefore, we have the following bound

$$
\exp\left\{\frac{n}{2}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2\left(Z_{ij}^{(a)}\left(Z_{ij}^{(a)^T}Z_{ij}^{(a)}\right)^{-1}\right)^2\right)\right\} \leq \exp\left\{\frac{1}{2n^{1-2\alpha}}\left(\sum_{i=k_0+1}^{k_1}\sum_{j=k_0+1}^{k_1}\lambda_1^2\left(Z_{ij}^{(a)}\right)^2\right)\right\}
$$

under $\left(E_{n,\alpha}^{(a)}\right)^C$. This term is $\mathcal{O}_{\mathbb{P}}(1)$, and from Lemma 5, we know that $\mathbb{1}_{\left(E_{n,\alpha}^{(a)}\right)^C}$ is $o_{\mathbb{P}}(1)$, therefore, this second term is $o_{\mathbb{P}}(1)$.

Now, if $\lambda_1(k_1 - k_0) = \mathcal{O}(1)$, then as n goes to infinity, this term converges to

$$
\exp\left\{\lim_{n\to\infty}\lambda_1^2\frac{(k_1-k_0)^2}{2}\right\} < \infty.
$$

Thus, under this condition, the null and alternate hypotheses are mutually contiguous. Due to the optimality of the likelihood ratio test, this shows that no test is asymptotically powerful. Additionally, we can note that if $\lambda_1 (k_1 - k_0) = o(1)$, then the second moment of the likelihood ratio converges in expectation to 1, which means the likelihood ratio test is asymptotically powerless, and in turn, all tests are asymptotically powerless.

Appendix F. Proof of Theorem 3

We will assume the general case, with $k_1 > k_0, \lambda_0 > \lambda_1$, and find the regimes under which each test has asymptotic consistency.

F.1. The trace test

First consider the trace test $\mathcal{T}_1$. By stochastic monotonicity, it is sufficient to consider all primary spikes and all secondary spikes being homogeneous, respectively. Therefore, we have

$$\begin{aligned}\mathrm{Tr}(X) &= \mathrm{Tr}(U\Lambda U^T) + \mathrm{Tr}(W)\\ &= k_0\lambda_0 + (k_1 - k_0)\lambda_1 + \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \sqrt{n} W_{jj}.\end{aligned}$$

Since $\sqrt{n}W_{jj}, 1 \le j \le n$, are i.i.d. zero mean random variables with variance 2, it follows from the CLT that

$$\mathrm{Tr}(X) - k_0\lambda_0 + (k_1 - k_0)\lambda_1 \xrightarrow{d} \mathcal{N}(0, 2).$$

This can be used to construct an asymptotic size-α test. In fact, if $\frac{1}{\sqrt{n}}W$ is a GOE matrix then $\mathrm{Tr}(X) - k_0\lambda_0 + (k_1 - k_0)\lambda_1 \sim \mathcal{N}(0, 2)$ and the constructed test will have exact size α.

Under H_0, $\lambda_1 = 0$ so that

$$\mathrm{Tr}(X) - k_0\lambda_0 \xrightarrow{d} \mathcal{N}(0, 2).$$

Therefore, to achieve an asymptotic size of α we define the test as follows:

$$\mathcal{T}_1 : \text{Reject } H_0 \text{ if } \mathrm{Tr}(X) > k_0\lambda_0 + \sqrt{2}z_\alpha.$$

As discussed above, this is of exact size-α for GOE. The power of this test is

$$\begin{aligned}\mathbb{P}_{H_1}(\mathrm{Tr}(X) > k_0\lambda_0 + \sqrt{2}z_\alpha) &= \mathbb{P}_{H_1}\left(\frac{1}{\sqrt{2}}(\mathrm{Tr}(X) - k_0\lambda_0 - (k_1 - k_0)\lambda_1) > z_\alpha - (k_1 - k_0)\frac{\lambda_1}{\sqrt{2}}\right)\\ &= \bar{\Phi}\left(z_\alpha - \frac{(k_1 - k_0)\lambda_1}{\sqrt{2}}\right) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\\ &= \Phi\left(\frac{(k_1 - k_0)\lambda_1}{\sqrt{2}} - z_\alpha\right) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

where the second line follows from the Berry-Esseen CLT and $\bar{\Phi}(x) = 1 - \Phi(x)$.

Therefore, for the trace test to be asymptotically powerful, we need $(k_1 - k_0)\lambda_1 \to \infty$. From the Mill's ratio inequalities, we have the following asymptotics on the Gaussian CDF Φ: As $x \to \infty$,

$$\bar{\Phi}(x) \sim \frac{1}{\sqrt{2\pi}x} \exp\left\{-x^2/2\right\}.$$

Thus, if $(k_1 - k_0)\lambda_1 \to \infty$, we have

$$\mathbb{P}_{H_1}(\mathcal{T}_1 \text{ is rejected}) = 1 - \mathcal{O}\left(\exp\left\{-\Theta\left((k_1-k_0)^2\lambda_1^2\right)\right\}\right) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

F.2. The partial trace test

We denote the partial trace test of the order k as $\mathrm{PT}_k(X)$. We will use the result from Benaych-Georges et al. (2011) to find the phase transitions of the eigenvalues. Once again by stochastic monotonicity, it is sufficient to consider all primary spikes and all secondary spikes being homogeneous, respectively.

Since our base spectral measure for the random matrix would correspond to the semicircle distribution, we can use Theorem 1.3 and 1.4 from Benaych-Georges et al. (2011)

Define a function $\varpi : [0,\infty) \to [2,\infty)$ as follows:

$$\varpi(x) = \begin{cases} x + \frac{1}{x} & \text{if } x > 1, \\ 2 & \text{if } 0 \le x \le 1. \end{cases}$$

We will be using a value of k that is comparable to k_1. For the largest eigenvalues, we have the following result, which arises from the BBP transition described in Baik et al. (2004):

$$\begin{aligned} &\text{If } i \le k_1, \qquad \text{then } \eta_i \xrightarrow{\text{a.s.}} \varpi(\lambda_i), \qquad \text{for } \lambda_1 > 1 \\ &\text{If } i > k_1, \qquad \text{then } \eta_i \xrightarrow{\text{a.s.}} 2, \end{aligned}$$

where η_i is the i-th largest eigenvalue of X, and λ_i is the i-th largest eigenvalue of Λ. We also have the following fluctuation result due to Benaych-Georges et al. (2011). See also Theorem 2.14 of Knowles and Yin (2013) and Theorem 2.11 of Knowles and Yin (2014).

Theorem 6 (Benaych-Georges et al. (2011, Theorem 3.2)) *Let α_j be the distinct values of $\lambda_i > 1$. Let q denote the total number of such values, and q_j denote the number of repetitions of α_j. Then the joint asymptotic distribution of $\left(\sqrt{n}\left(\eta_i - \varpi(\lambda_i)\right)\right)_{1\le i\le \sum_j q_j}$ is the same as the joint distribution of the eigenvalues of $(C_{\alpha_j} M_j)_{1\le j\le q}$, where $C_\alpha = \sqrt{\frac{\alpha^2-1}{\alpha}}$, M_j's are independent, unscaled $GOE(q_j)$ random matrices. On the other hand, the non-critical eigenvalues have Tracy-Widom fluctuations*

$$n^{2/3}(\eta_i - 2) \xrightarrow{d} \mathrm{TW}_i,$$

where TW_i is the i-th order Tracy-Widom law.

Consider a fixed $k > k_1$. If $\lambda_1 > 1$, then we have k_1 critical and $k - k_1$ non-critical eigenvalues contributing to $\mathrm{PT}_k(X)$. Let Z_0, Z_1 be i.i.d. standard Gaussians. By Theorem 6,

$$\begin{aligned} \sqrt{n}\Big(\mathrm{PT}_k(X) - k_0\varpi(\lambda_0) - &(k_1-k_0)\varpi(\lambda_1) - 2(k-k_1)\Big) \\ &\xrightarrow{d} C_{\lambda_0}\sqrt{2k_0}Z_0 + C_{\lambda_1}\sqrt{2(k_1-k_0)}Z_1 \\ &\stackrel{d}{=} N\left(0, 2k_0\frac{\lambda_0^2-1}{\lambda_0} + 2(k_1-k_0)\frac{\lambda_1^2-1}{\lambda_1}\right). \end{aligned}$$

Under H_0, $\lambda_1 = 0$ and therefore

$$\sqrt{n}\left(\mathrm{PT}_k(X) - k_0\varpi(\lambda_0) - 2(k-k_0)\right) \xrightarrow{d} N\left(0, 2k_0\frac{\lambda_0^2-1}{\lambda_0}\right).$$

(Note that we operate under the assumption that $\lambda_0 > 1$ so that the top k_0 eigenvalues are critical.) We choose $\lambda \geq 1$ as to the regime under which we will investigate the power function. Let

$$g(x) = k_0\left(x + \frac{1}{x}\right) + 2(k-k_0) + \sqrt{\frac{2k_0}{n}\frac{x^2-1}{x}}z_\alpha.$$

It is straightforward to check that g is strictly increasing on $(1,\infty)$. Since we have an upper bound on λ_0, namely β, we may now define the partial trace test as follows:

$$\mathcal{T}_2 : \text{Reject } H_0 \text{ if } \mathrm{PT}_k(X) > g(\beta).$$

Then, under H_0,

$$\mathbb{P}_{H_0}(\mathcal{T}_2 \text{ is rejected}) = \mathbb{P}_{H_0}(\mathrm{PT}_k(X) > g(\beta)) \leq \mathbb{P}_{H_0}(PT_k(X) > g(\lambda_0)) = \alpha + o(1).$$

This gives an aysmptotic level-α test. We now calculate the power for a given value of λ_1.

$$\begin{aligned}
&\mathbb{P}_{H_1}(\mathrm{PT}_k(X) > g(\beta)) \\
&= \mathbb{P}_{H_1}\left(\sqrt{n}\left(\mathrm{PT}_k(X) - k_0\varpi(\lambda_0) - (k_1-k_0)\varpi(\lambda_1) - 2(k-k_1)\right)\right. \\
&> \left.\sqrt{n}\left(k_0\varpi(\beta) + 2(k_1-k_0) + \sqrt{\frac{2k_0}{n}\frac{\beta^2-1}{\beta}}z_\alpha - k_0\varpi(\lambda_0) - (k_1-k_0)\varpi(\lambda_1)\right)\right) \\
&= \bar{\Phi}\left(\frac{\sqrt{n}\left(k_0\varpi(\beta) + 2(k_1-k_0) - k_0\varpi(\lambda_0) - (k_1-k_0)\varpi(\lambda_1)\right) + \sqrt{2k_0\frac{\beta^2-1}{\beta}}z_\alpha}{\sqrt{2k_0\frac{\lambda_0^2-1}{\lambda_0} + 2(k_1-k_0)\frac{\lambda_1^2-1}{\lambda_1}}}\right) + o(1) \\
&= \bar{\Phi}\left(\underbrace{\frac{\sqrt{n}\left(k_0(\varpi(\beta)-\varpi(\lambda_0)) - (k_1-k_0)\frac{(\lambda_1-1)^2}{\lambda_1}\right)}{\sqrt{2k_0\frac{\lambda_0^2-1}{\lambda_0} + 2(k_1-k_0)\frac{\lambda_1^2-1}{\lambda_1}}}}_{=:\mathrm{I}_{\mathrm{PT}}} + \underbrace{\sqrt{\frac{k_0\frac{\beta^2-1}{\beta}z_\alpha}{k_0\frac{\lambda_0^2-1}{\lambda_0} + (k_1-k_0)\frac{\lambda_1^2-1}{\lambda_1}}}}_{=:\mathrm{II}_{\mathrm{PT}}}\right) + o(1).
\end{aligned}$$

Observe that

$$\mathrm{II}_{\mathrm{PT}} = \sqrt{\frac{k_0\frac{\beta^2-1}{\beta}z_\alpha}{k_0\frac{\lambda_0^2-1}{\lambda_0} + (k_1-k_0)\frac{\lambda_1^2-1}{\lambda_1}}} < \sqrt{\frac{k_0\frac{\beta^2-1}{\beta}z_\alpha}{k_0\frac{\lambda_0^2-1}{\lambda_0}}} < \sqrt{\frac{\frac{\beta^2-1}{\beta}z_\alpha}{\frac{\lambda_0^2-1}{\lambda_0}}} < \infty$$

for $1 < \lambda_1 < \lambda_0 < \beta$. Thus, in order for $\mathcal{T}_2$ to be asymptotically powerful, we need $\mathrm{I}_{\mathrm{PT}} \to -\infty$ Towards that end, note that

$$-\mathrm{I}_{\mathrm{PT}} = \frac{\sqrt{n}\left((k_1-k_0)\frac{(\lambda_1-1)^2}{\lambda_1} - k_0(\varpi(\beta)-\varpi(\lambda_0))\right)}{\sqrt{2k_0\frac{\lambda_0^2-1}{\lambda_0} + 2(k_1-k_0)\frac{\lambda_1^2-1}{\lambda_1}}}.$$

Thus, the partial trace test $\mathcal{T}_2$ is asymptotically powerful if

$$\omega := (k_1 - k_0)\frac{(\lambda_1 - 1)^2}{\lambda_1} - k_0(\varpi(\beta) - \varpi(\lambda_0)) > 0.$$

Using the asymptotic rate of Φ, we have

$$\mathbb{P}_{H_1}(\mathcal{T}_2 \text{ is rejected}) = 1 - \mathcal{O}\left(\exp\left\{-\Theta\left(n\omega^2\right)\right\}\right) + o(1).$$

F.3. The largest non-critical eigenvalue test

Now, we will look at a test using the (k_0+1)-th sample eigenvalue η_{k_0+1}. Using the same result as before, we know that under the null, η_{k_0+1} is non-critical and has Tracy-Widom fluctuations:

$$n^{2/3}\left(\eta_{k_0+1} - 2\right) \xrightarrow{d} \mathrm{TW}_{k_0+1},$$

where TW_{k_0+1} is the (k_0+1)-th order Tracy-Widom law. Denoting by t_{k_0+1} the upper quantile function of TW_{k_0+1}, we may construct an asymptotic size-α test as follows:

$$\mathcal{T}_3 : \text{Reject } H_0 \text{ if } \eta_{k_0+1} > 2 + n^{-2/3} t_{k_0+1}(\alpha).$$

Under the alternative, we have

$$\sqrt{n}\left(\eta_{k_0+1} - \varpi(\lambda_1)\right) \xrightarrow{d} N\left(0, 2\frac{\lambda_1^2 - 1}{\lambda_1}\right).$$

The power of $\mathcal{T}_3$ under this setup is

$$\begin{aligned}
&\mathbb{P}_{H_1}\left(\eta_{k_0+1} > 2 + n^{-2/3}t_{k_0+1}(\alpha)\right) \\
&\quad= \mathbb{P}_{H_1}\left(\sqrt{n}\left(\eta_{k_0+1} - \varpi(\lambda_1)\right) > \sqrt{n}\left(2 + n^{-2/3}t_{k_0+1}(\alpha) - \varpi(\lambda_1)\right)\right) \\
&\quad= \bar{\Phi}\left(\frac{\sqrt{n}\left(2 + n^{-2/3}t_{k_0+1}(\alpha) - \left(\lambda_1 + \frac{1}{\lambda_1}\right)\right)}{\sqrt{2\frac{\lambda_1^2-1}{\lambda_1}}}\right) + o(1) \\
&\quad= \bar{\Phi}\left(\frac{\sqrt{n}\left(-\left(\sqrt{\lambda_1} - \frac{1}{\sqrt{\lambda_1}}\right)^2 + n^{-2/3}t_{k_0+1}(\alpha)\right)}{\sqrt{2\left(\lambda_1 - \frac{1}{\lambda_1}\right)}}\right) + o(1) \\
&\quad= \bar{\Phi}\left(\underbrace{\frac{-\sqrt{n}\left(\sqrt{\lambda_1} - \frac{1}{\sqrt{\lambda_1}}\right)^2}{\sqrt{2\left(\lambda_1 - \frac{1}{\lambda_1}\right)}}}_{=:\mathrm{I}_{\lambda_1}} + \underbrace{\frac{n^{-1/6}t_{k_0+1}(\alpha)}{\sqrt{2\left(\lambda_1 - \frac{1}{\lambda_1}\right)}}}_{=:\mathrm{II}_{\lambda_1}}\right) + o(1),
\end{aligned}$$

where the second equality follows from Pólya's theorem (see, e.g., Lemma 2.11 in Van der Vaart (2000)). Since $\beta \geq \lambda_1 > 1$, it is clear that $\mathrm{II}_{\lambda_1} = o(1)$. Therefore $\mathcal{T}_3$ is asymptotically powerful if $\mathrm{I}_{\lambda_1} \to -\infty$. Indeed,

$$\mathrm{I}_{\lambda_1} = \frac{-\sqrt{n}\left(\sqrt{\lambda_1} - \frac{1}{\sqrt{\lambda_1}}\right)^2}{\sqrt{2\left(\lambda_1 - \frac{1}{\lambda_1}\right)}} = \frac{-\sqrt{n}\,(\lambda_1 - 1)^{3/2}}{\sqrt{2\lambda_1(1+\lambda_1)}} \leq \frac{-\sqrt{n}\,(\lambda_1 - 1)^{3/2}}{\sqrt{2\beta(1+\beta)}} \to -\infty,$$

under the assumption that $\beta \geq \lambda_1 > 1$. In fact, the power function has the following asymptotics:

$$\mathbb{P}_{H_1}(\mathcal{T}_3 \text{ is rejected}) = 1 - \mathcal{O}\left(\exp\left\{-\Theta\left(n\,(\lambda_1 - 1)^3\right)\right\}\right) + o(1).$$

Remark *In fact, using the rigidity results of Knowles and Yin (2013) and Knowles and Yin (2014), it is possible to obtain these results under the assumption* $\lambda_1 - 1 = \Omega((\log n)^{C \log\log n} n^{-1/3})$.