On the Limitation of Spectral Methods: From the Gaussian Hidden Clique Problem to Rank-One Perturbations of Gaussian Tensors

Andrea Montanari
Department of Electrical Engineering and Department of Statistics. Stanford University.
montanari@stanford.edu

Daniel Reichman
Department of Cognitive and Brain Sciences, University of California, Berkeley, CA
daniel.reichman@gmail.com

Ofer Zeitouni
Faculty of Mathematics, Weizmann Institute, Rehovot 76100, Israel
and Courant Institute, New York University
ofer.zeitouni@weizmann.ac.il

Abstract

We consider the following detection problem: given a realization of a symmetric matrix $\mathbf{X}$ of dimension n, distinguish between the hypothesis that all upper triangular variables are i.i.d. Gaussians variables with mean 0 and variance 1 and the hypothesis that there is a planted principal submatrix $\mathbf{B}$ of dimension L for which all upper triangular variables are i.i.d. Gaussians with mean 1 and variance 1, whereas all other upper triangular elements of $\mathbf{X}$ not in $\mathbf{B}$ are i.i.d. Gaussians variables with mean 0 and variance 1. We refer to this as the 'Gaussian hidden clique problem'. When $L = (1+\epsilon)\sqrt{n}$ ($\epsilon > 0$), it is possible to solve this detection problem with probability $1 - o_n(1)$ by computing the spectrum of $\mathbf{X}$ and considering the largest eigenvalue of $\mathbf{X}$. We prove that when $L < (1-\epsilon)\sqrt{n}$ no algorithm that examines only the eigenvalues of $\mathbf{X}$ can detect the existence of a hidden Gaussian clique, with error probability vanishing as $n \to \infty$. The result above is an immediate consequence of a more general result on rank-one perturbations of k-dimensional Gaussian tensors. In this context we establish a lower bound on the critical signal-to-noise ratio below which a rank-one signal cannot be detected.

1 Introduction

Consider the following detection problem. One is given a symmetric matrix $\mathbf{X} = \mathbf{X}(n)$ of dimension n, such that the $\binom{n}{2} + n$ entries $(\mathbf{X}_{i,j})_{i \le j}$ are *mutually independent* random variables. Given (a realization of) $\mathbf{X}$ one would like to distinguish between the hypothesis that all random variables $\mathbf{X}_{i,j}$ have the same distribution F_0 to the hypothesis where there is a set $U \subseteq [n]$, with $L := |U|$, so that all random variables in the submatrix $\mathbf{X}_U := (\mathbf{X}_{s,t} : s, t \in U)$ have a distribution F_1 that is different from the distribution of all other elements in $\mathbf{X}$ which are still distributed as F_0. We refer to $\mathbf{X}_U$ as the *hidden submatrix*.

The same problem was recently studied in [1, 8] and, for the asymmetric case (where no symmetry assumption is imposed on the independent entries of $\mathbf{X}$), in [6, 18, 20]. Detection problems with similar flavor (such as the hidden clique problem) have been studied over the years in several fields including computer science, physics and statistics. We refer to Section 5 for further discussion of the related literature. An intriguing outcome of these works is that, while the two hypothesis are statistically distinguishable as soon as $L \geq C \log n$ (for C a sufficiently large constant) [7], practical algorithms require significantly larger L. In this paper we study the class of spectral (or eigenvalue-based) tests detecting the hidden submatrix. Our proof technique naturally allow to consider two further generalizations of this problem that are of independent interests. We briefly summarize our results below.

The Gaussian hidden clique problem. This is a special case of the above hypothesis testing setting, whereby $F_0 = \mathsf{N}(0, 1)$ and $F_1 = \mathsf{N}(1, 1)$ (entries on the diagonal are defined slightly differently in order to simplify calculations). Here and below $\mathsf{N}(m, \sigma^2)$ denote the Gaussian distribution of mean m and variance σ^2. Equivalently, let $\mathbf{Z}$ be a random matrix from the Gaussian Orthogonal Ensemble (GOE) i.e. $\mathbf{Z}_{ij} \sim \mathsf{N}(0, 1/n)$ independently for $i < j$, and $\mathbf{Z}_{ii} \sim \mathsf{N}(0, 2/n)$. Then, under hypothesis $H_{1,L}$ we have $\mathbf{X} = n^{-1/2}\mathbf{1}_U\mathbf{1}_U^{\mathsf{T}} + \mathbf{Z}$ ($\mathbf{1}_U$ being the indicator vector of U), and under hypothesis H_0, $\mathbf{X} = \mathbf{Z}$ (the factor n in the normalization is for technical convenience). The Gaussian hidden clique problem can be thought of as the following clustering problem: there are n elements and the entry (i, j) measures the similarity between elements i and j. The hidden submatrix corresponds to a cluster of similar elements, and our goal is to determine given the matrix whether there is a large cluster of similar elements or alternatively, whether all similarities are essentially random (Gaussian) noise.

Our focus in this work is on the following restricted hypothesis testing question. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the ordered eigenvalues of $\mathbf{X}$. *Is there a test that depends only on $\lambda_1, \ldots, \lambda_n$ and that distinguishes H_0 from $H_{1,L}$ 'reliably,' i.e. with error probability converging to 0 as $n \to \infty$?* Notice that the eigenvalues distribution does not depend on U as long as this is independent from the noise $\mathbf{Z}$. We can therefore think of U as fixed for this question. Historically, the first polynomial time algorithm for detecting a planted clique of size $O(\sqrt{n})$ in a random graph [2] relied on spectral methods (see Section 5 for more details). This is one reason for our interest in spectral tests for the Gaussian hidden clique problem.

If $L \geq (1 + \varepsilon)\sqrt{n}$ then [11] implies that a simple test checking whether $\lambda_1 \geq 2 + \delta$ for some $\delta = \delta(\varepsilon) > 0$ is reliable for the Gaussian hidden clique problem. We prove that this result is tight, in the sense that no spectral test is reliable for $L \leq (1 - \varepsilon)\sqrt{n}$.

Rank-one matrices in Gaussian noise. Our proof technique builds on a simple observation. Since the noise $\mathbf{Z}$ is invariant under orthogonal transformations[1], the above question is equivalent to the following testing problem. For $\beta \in \mathbb{R}_{\geq 0}$, and $\mathbf{v} \in \mathbb{R}^n$, $\|\mathbf{v}\|_2 = 1$ a uniformly random unit vector, test H_0: $\mathbf{X} = \mathbf{Z}$ versus H_1, $\mathbf{X} = \beta\mathbf{v}\mathbf{v}^{\mathsf{T}} + \mathbf{Z}$. (The correspondence between the two problems yields $\beta = L/\sqrt{n}$.)

Again, this problem (and a closely related asymmetric version [22]) has been studied in the literature, and it follows from [11] that a reliable test exists for $\beta \geq 1 + \varepsilon$. We provide a simple proof (based on the second moment method) that no test is reliable for $\beta < 1 - \varepsilon$.

Rank-one tensors in Gaussian noise. It turns that the same proof applies to an even more general problem: detecting a rank-one signal in a noisy tensor. We carry out our analysis in this more general setting for two reasons. First, we think that this clarifies the what aspects of the model are important for our proof technique to apply. Second, the problem estimating tensors from noisy data has attracted significant interest recently within the machine learning community [15, 21].

More precisely, we consider a noisy tensor $\mathbf{X} \in \bigotimes^k \mathbb{R}^n$, of the form $\mathbf{X} = \beta\, \mathbf{v}^{\otimes k} + \mathbf{Z}$, where $\mathbf{Z}$ is Gaussian noise, and $\mathbf{v}$ is a random unit vector. We consider the problem of testing this hypothesis against H_0: $\mathbf{X} = \mathbf{Z}$. We establish a threshold $\beta_k^{\rm 2nd}$ such that no test can be reliable for $\beta < \beta_k^{\rm 2nd}$ (in particular $\beta_2^{\rm 2nd} = 1$). Two differences are worth remarking for $k \geq 3$ with respect to the more familiar matrix case $k = 2$. First, we do not expect the second moment bound $\beta_k^{\rm 2nd}$ to be tight, i.e. a reliable test to exist for all $\beta > \beta_k^{\rm 2nd}$. On the other hand, we can show that it is tight up to

[1]By this we mean that, for any orthogonal matrix $\mathbf{R} \in O(n)$, independent of $\mathbf{Z}$, $\mathbf{R}\mathbf{Z}\mathbf{R}^{\mathsf{T}}$ is distributed as $\mathbf{Z}$.

a universal (k and n independent) constant. Second, below $\beta_k^{\rm 2nd}$ the problem is more difficult than the matrix version below $\beta_2^{\rm 2nd} = 1$: not only no reliable test exists but, asymptotically, any test behaves asymptotically as random guessing. For more details on our results regarding noisy tensors, see Theorem 3.

2 Main result for spectral detection

Let $\mathbf{Z}$ be a GOE matrix as defined in the previous section. Equivalently if $\mathbf{G}$ is an (asymmetric) matrix with i.i.d. entries $\mathbf{G}_{i,j} \sim \mathsf{N}(0,1)$,

$$\mathbf{Z} = \frac{1}{\sqrt{2n}}\big(\mathbf{G} + \mathbf{G}^{\mathsf{T}}\big)\,. \tag{1}$$

For a deterministic sequence of vectors $\mathbf{v}(n)$, $\|\mathbf{v}(n)\|_2 = 1$, we consider the two hypotheses

$$\begin{cases} H_0: & \mathbf{X} = \mathbf{Z}\,,\\ H_{1,\beta}: & \mathbf{X} = \beta\mathbf{v}\mathbf{v}^{\mathsf{T}} + \mathbf{Z}\,. \end{cases} \tag{2}$$

A special example is provided by the Gaussian hidden clique problem in which case $\beta = L/\sqrt{n}$ and $\mathbf{v} = \mathbf{1}_U/\sqrt{L}$ for some set $U \subseteq [n]$, $|U| = L$,

$$\begin{cases} H_0: & \mathbf{X} = \mathbf{Z}\,,\\ H_{1,L}: & \mathbf{X} = \frac{1}{\sqrt{n}}\mathbf{1}_U\mathbf{1}_U^{\mathsf{T}} + \mathbf{Z}\,. \end{cases} \tag{3}$$

Observe that the distribution of eigenvalues of $\mathbf{X}$, under either alternative, is invariant to the choice of the vector $\mathbf{v}$ (or subset U), as long as the norm of $\mathbf{v}$ is kept fixed. Therefore, any successful algorithm that examines only the eigenvalues, will distinguish between H_0 and $H_{1,\beta}$ but not give any information on the vector $\mathbf{v}$ (or subset U, in the case of $H_{1,L}$).

We let $Q_0 = Q_0(n)$ (respectively, $Q_1 = Q_1(n)$) denote the distribution of the eigenvalues of $\mathbf{X}$ under H_0 (respectively $H_1 = H_{1,\beta}$ or $H_{1,L}$).

A *spectral statistical test* for distinguishing between H_0 and H_1 (or simply a spectral test) is a measurable map $T_n : (\lambda_1, \dots, \lambda_n) \mapsto \{0, 1\}$. To formulate precisely what we mean by the word *distinguish*, we introduce the following notion.

Definition 1. *For each $n \in \mathbb{N}$, let $\mathbb{P}_{0,n}$, $\mathbb{P}_{1,n}$ be two probability measures on the same measure space $(\Omega_n, \mathcal{F}_n)$. We say that the sequence $(\mathbb{P}_{1,n})$ is contiguous with respect to $(\mathbb{P}_{0,n})$ if, for any sequence of events $A_n \in \mathcal{F}_n$,*

$$\lim_{n\to\infty} \mathbb{P}_{0,n}(A_n) = 0 \;\;\Rightarrow\;\; \lim_{n\to\infty} \mathbb{P}_{1,n}(A_n) = 0\,. \tag{4}$$

Note that contiguity is not in general a symmetric relation.

In the context of the spectral statistical tests described above, the sequences A_n in Definition 1 (with $P_n = Q_0(n)$ and $Q_n = Q_1(n)$) can be put in correspondence with spectral statistical tests T_n by taking $A_n = \{(\lambda_1, \dots, \lambda_n) : T_n(\lambda_1, \dots, \lambda_n) = 0\}$. We will thus say that H_1 is *spectrally contiguous* with respect to H_0 if Q_n is contiguous with respect to P_n.

Our main result on the Gaussian hidden clique problem is the following.

Theorem 1. *For any sequence $L = L(n)$ satisfying $\limsup_{n\to\infty} L(n)/\sqrt{n} < 1$, the hypotheses $H_{1,L}$ are spectrally contiguous with respect to H_0.*

2.1 Contiguity and integrability

Contiguity is related to a notion of uniform absolute continuity of measures. Recall that a probability measure μ on a measure space is *absolutely continuous* with respect to another probability measure ν if for every measurable set A, $\nu(A) = 0$ implies that $\mu(A) = 0$, in which case there exists a ν-integrable, non-negative function $f \equiv \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ (the *Radon-Nikodym derivative* of μ with respect to ν), so that $\mu(A) = \int_A f\,\mathrm{d}\nu$ for every measurable set A. We then have the following known useful fact:

Lemma 2. *Within the setting of Definition 1, assume that $\mathbb{P}_{1,n}$ is absolutely continuous with respect to $\mathbb{P}_{0,n}$, and denote by $\Lambda_n \equiv \frac{\mathrm{d}\mathbb{P}_{1,n}}{\mathrm{d}\mathbb{P}_{0,n}}$ its Radon-Nikodym derivative.*

(a) If $\limsup_{n\to\infty} \mathbb{E}_{0,n}(\Lambda_n^2) < \infty$, then $(\mathbb{P}_{1,n})$ is contiguous with respect to $(\mathbb{P}_{0,n})$.
(b) If $\lim_{n\to\infty} \mathbb{E}_{0,n}(\Lambda_n^2) = 1$, then $\lim_{n\to\infty} \|\mathbb{P}_{0,n} - \mathbb{P}_{1,n}\|_{\mathrm{TV}} = 0$, where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation distance, i.e.

$$\|\mathbb{P}_{0,n} - \mathbb{P}_{1,n}\|_{\mathrm{TV}} \equiv \sup_A |\mathbb{P}_{0,n}(A) - \mathbb{P}_{1,n}(A)\|.$$

2.2 Method and structure of the paper

Consider problem (2). We use the fact that the law of the eigenvalues under both H_0 and $H_{1,\beta}$ are invariant under conjugations by a orthogonal matrix. Once we conjugate matrices sampled under the hypothesis $H_{1,\beta}$ by an independent orthogonal matrix sampled according to the Haar distribution, we get a matrix distributed as

$$\mathbf{X} = \beta \mathbf{v}\mathbf{v}^{\mathsf{T}} + \mathbf{Z}\,, \tag{5}$$

where $\mathbf{u}$ is uniform on the n-dimensional sphere, and $\mathbf{Z}$ is a GOE matrix (with off-diagonal entries of variance $1/n$). Letting $\mathbb{P}_{1,n}$ denote the law of $\beta\mathbf{u}\mathbf{u}^{\mathsf{T}} + \mathbf{Z}$ and $\mathbb{P}_{0,n}$ denote the law of $\mathbf{Z}$, we show that $\mathbb{P}_{1,n}$ is contiguous with respect to $\mathbb{P}_{0,n}$, which implies that the law of eigenvalues $Q_1(n)$ is contiguous with respect to $Q_0(n)$.

To show the contiguity, we consider a more general setup, of independent interest, of Gaussian tensors of order k, and in that setup show that the Radon-Nikodym derivative $\Lambda_{n,L} = \frac{\mathrm{d}\mathbb{P}_{1,n}}{\mathrm{d}\mathbb{P}_{0,n}}$ is uniformly square integrable under $\mathbb{P}_{0,n}$; an application of Lemma 2 then quickly yields Theorem 1.

The structure of the paper is as follows. In the next section, we define formally the detection problem for a symmetric tensor of order $k \geq 2$. We show the existence of a threshold under which detection is not possible (Theorem 3), and show how Theorem 1 follows from this. Section 4 is devoted to the proof of Theorem 3, and concludes with some additional remarks and consequences of Theorem 3. Finally, Section 5 is devoted to a description of the relation between the Gaussian hidden clique problem and hidden clique problem in computer science, and related literature.

3 A symmetric tensor model and a reduction

Exploiting rotational invariance, we will reduce the spectral detection problem to a detection problem involving a standard detection problem between random matrices. Since the latter generalizes to a tensor setup, we first introduce a general Gaussian hypothesis testing for k-tensors, which is of independent interest. We then explain how the spectral detection problem reduces to the special case of $k = 2$.

3.1 Preliminaries and notation

We use lower-case boldface for vectors (e.g. $\mathbf{u}$, $\mathbf{v}$) and upper-case boldface for matrices and tensors (e.g. $\mathbf{X}, \mathbf{Z}$). The ordinary scalar product and ℓ_p norm over vectors are denoted by $\langle \mathbf{u}, \mathbf{v}\rangle = \sum_{i=1}^n \mathbf{u}_i \mathbf{v}_i$, and $\|\mathbf{v}\|_p$. We write $\mathbb{S}^{n-1}$ for the unit sphere in n dimensions

$$\mathbb{S}^{n-1} \equiv \left\{\mathbf{x} \in \mathbb{R}^n :\ \ \|\mathbf{x}\|_2 = 1\right\}. \tag{6}$$

Given $\mathbf{X} \in \bigotimes^k \mathbb{R}^n$ a real k-th order tensor, we let $\{\mathbf{X}_{i_1,\dots,i_k}\}_{i_1,\dots,i_k}$ denote its coordinates. The outer product of two tensors is $\mathbf{X} \otimes \mathbf{Y}$, and, for $\mathbf{v} \in \mathbb{R}^n$, we define $\mathbf{v}^{\otimes k} = \mathbf{v} \otimes \cdots \otimes \mathbf{v} \in \bigotimes^k \mathbb{R}^n$ as the k-th outer power of $\mathbf{v}$. We define the inner product of two tensors $\mathbf{X}, \mathbf{Y} \in \bigotimes^k \mathbb{R}^n$ as

$$\langle \mathbf{X}, \mathbf{Y}\rangle = \sum_{i_1,\cdots,i_k \in [n]} \mathbf{X}_{i_1,\cdots,i_k} \mathbf{Y}_{i_1,\cdots,i_k}\ . \tag{7}$$

We define the Frobenius (Euclidean) norm of a tensor $\mathbf{X}$ by $\|\mathbf{X}\|_F = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle}$, and its operator norm by

$$\|\mathbf{X}\|_{op} \equiv \max\{\langle \mathbf{X}, \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_k \rangle \ : \ \forall i \in [k] \ , \ \|\mathbf{u}_i\|_2 \le 1\}. \tag{8}$$

It is easy to check that this is indeed a norm. For the special case $k = 2$, it reduces to the ordinary ℓ_2 matrix operator norm (equivalently, to the largest singular value of $\mathbf{X}$).

For a permutation $\pi \in \mathfrak{S}_k$, we will denote by $\mathbf{X}^\pi$ the tensor with permuted indices $\mathbf{X}^\pi_{i_1,\cdots,i_k} = \mathbf{X}_{\pi(i_1),\cdots,\pi(i_k)}$. We call the tensor $\mathbf{X}$ *symmetric* if, for any permutation $\pi \in \mathfrak{S}_k$, $\mathbf{X}^\pi = \mathbf{X}$. It is proved [23] that, for symmetric tensors, we have the equivalent representation

$$\|\mathbf{X}\|_{op} \equiv \max\{|\langle \mathbf{X}, \mathbf{u}^{\otimes k} \rangle| \ : \ \ \|\mathbf{u}\|_2 \le 1\}. \tag{9}$$

We define $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \infty$ with the usual conventions of arithmetic operations.

3.2 The symmetric tensor model and main result

We denote by $\mathbf{G} \in \bigotimes^k \mathbb{R}^n$ a tensor with independent and identically distributed entries $\mathbf{G}_{i_1,\cdots,i_k} \sim \mathsf{N}(0,1)$ (note that this tensor is not symmetric).

We define the *symmetric standard normal* noise tensor $\mathbf{Z} \in \bigotimes^k \mathbb{R}^n$ by

$$\mathbf{Z} = \frac{1}{k!}\sqrt{\frac{2}{n}} \sum_{\pi \in \mathfrak{S}_k} \mathbf{G}^\pi \, . \tag{10}$$

Note that the subset of entries with unequal indices form an i.i.d. collection $\{\mathbf{Z}_{i_1,i_2,\dots,i_k}\}_{i_1<\cdots<i_k} \sim \mathsf{N}(0, 2/(n(k!)))$.

With this normalization, we have, for any symmetric tensor $\mathbf{A} \in \bigotimes^k \mathbb{R}^n$

$$\mathbb{E}\big\{e^{\langle \mathbf{A}, \mathbf{Z} \rangle}\big\} = \exp\Big\{\frac{1}{n}\|\mathbf{A}\|_F^2\Big\} \, . \tag{11}$$

We will also use the fact that $\mathbf{Z}$ is invariant in distribution under conjugation by orthogonal transformations, that is, that for any orthogonal matrix $U \in O(n)$, $\{\mathbf{Z}_{i_1,\dots,i_k}\}$ has the same distribution as $\{\sum_{j_1,\dots,j_k}\Big(\prod_{\ell=1}^k U_{i_\ell,j_\ell}\Big) \cdot \mathbf{Z}_{j_1,\dots,j_k}\}$.

Given a parameter $\beta \in \mathbb{R}_{\ge 0}$, we consider the following model for a random symmetric tensor $\mathbf{X}$:

$$\mathbf{X} \equiv \beta \, \mathbf{v}^{\otimes k} + \mathbf{Z} \, , \tag{12}$$

with $\mathbf{Z}$ a standard normal tensor, and $\mathbf{v}$ uniformly distributed over the unit sphere $\mathbb{S}^{n-1}$. In the case $k = 2$ this is the standard rank-one deformation of a GOE matrix.

We let $\mathbb{P}_\beta = \mathbb{P}_\beta^{(k)}$ denote the law of $\mathbf{X}$ under model (12).

Theorem 3. *For $k \ge 2$, let*

$$\beta_k^{\rm 2nd} \equiv \inf_{q \in (0,1)} \sqrt{-\frac{1}{q^k}\log(1-q^2)} \, . \tag{13}$$

Assume $\beta < \beta_k^{\rm 2nd}$. Then, for any $k \ge 3$, we have

$$\lim_{n\to\infty} \big\|\mathbb{P}_\beta - \mathbb{P}_0\big\|_{\rm TV} = 0 \, . \tag{14}$$

Further, for $k = 2$ and $\beta < \beta_k^{\rm 2nd} = 1$, $\mathbb{P}_\beta$ is contiguous with respect to $\mathbb{P}_0$.

A few remarks are in order, following Theorem 3.

First, it is not difficult to derive the asymptotic $\beta_k^{\rm 2nd} = \sqrt{\log(k/2)} + o_k(1)$ for large k.

Second, for $k = 2$ we get using $\log(1 - q^2) \leq -q^2$, that $\beta_k^{\rm 2nd} = 1$. Recall that for $k = 2$ and $\beta > 1$, it is known that the largest eigenvalue of $\mathbf{X}$, $\lambda_1(\mathbf{X})$ converges almost surely to $(\beta + 1/\beta)$ [11]. As a consequence $\|\mathbb{P}_0 - \mathbb{P}_\beta\|_{\rm TV} \to 1$ for all $\beta > 1$: the second moment bound is tight.

For $k \geq 3$, it follows by the triangle inequality that $\|\mathbf{X}\|_{op} \geq \beta - \|\mathbf{Z}\|_{op}$, and further $\limsup_{n\to\infty} \|\mathbf{Z}\|_{op} \leq \mu_k$ almost surely as $n \to \infty$ [19, 5] for some bounded μ_k. It follows that $\|\mathbb{P}_0 - \mathbb{P}_\beta\|_{\rm TV} \to 1$ for all $\beta > 2\mu_k$ [21]. Hence, the second moment bound is off by a k-dependent factor. For large k, $2\mu_k = \sqrt{2\log k} + O_k(1)$ and hence the factor is indeed bounded in k.

Behavior below the threshold. Let us stress an important qualitative difference between $k = 2$ and $k \geq 3$, for $\beta < \beta_k^{\rm 2nd}$. For $k \geq 3$, the two models are indistinguishable and any test is essentially as good as random guessing. Formally, for any measurable function $T : \otimes^k\mathbb{R}^n \to \{0, 1\}$, we have

$$\lim_{n\to\infty} \big[\mathbb{P}_0(T(\mathbf{X}) = 1) + \mathbb{P}_\beta(T(\mathbf{X}) = 0)\big] = 1\,. \tag{15}$$

For $k = 2$, our result implies that, for $\beta < 1$, $\|\mathbb{P}_0 - \mathbb{P}_\beta\|_{\rm TV}$ is bounded away from 1. On the other hand, it is easy to see that it is bounded away from 0 as well, i.e.

$$0 < \liminf_{n\to\infty} \|\mathbb{P}_0 - \mathbb{P}_\beta\|_{\rm TV} \leq \limsup_{n\to\infty} \|\mathbb{P}_0 - \mathbb{P}_\beta\|_{\rm TV} < 1\,. \tag{16}$$

Indeed, consider for instance the statistics $S = {\rm Tr}(\mathbf{X})$. Under $\mathbb{P}_0$, $S \sim {\sf N}(0, 2)$, while under $\mathbb{P}_\beta$, $S \sim {\sf N}(\beta, 2)$. Hence

$$\liminf_{n\to\infty} \|\mathbb{P}_0 - \mathbb{P}_\beta\|_{\rm TV} \geq \|{\sf N}(0, 1) - {\sf N}(\beta/\sqrt{2}, 1)\|_{\rm TV} = 1 - 2\Phi\Big(-\frac{\beta}{2\sqrt{2}}\Big) > 0 \tag{17}$$

(Here $\Phi(x) = \int_{-\infty}^x e^{-z^2/2}{\rm d}z/\sqrt{2\pi}$ is the Gaussian distribution function.) The same phenomenon for rectangular matrices ($k = 2$) is discussed in detail in [22].

3.3 Reduction of spectral detection to the symmetric tensor model, $k = 2$

Recall that in the setup of Theorem 1, $Q_{0,n}$ is the law of the eigenvalues of $\mathbf{X}$ under H_0 and $Q_{1,n}$ is the law of the eigenvalues of $\mathbf{X}$ under $H_{1,L}$. Then $Q_{1,n}$ is invariant by conjugation of orthogonal matrices. Therefore, the detection problem is not changed if we replace $\mathbf{X} = n^{-1/2}\mathbf{1}_U\mathbf{1}_U^{\sf T} + \mathbf{Z}$ by

$$\widehat{\mathbf{X}} \equiv \mathbf{R}\mathbf{X}\mathbf{R}^{\sf T} = \frac{1}{\sqrt{n}}\mathbf{R}\mathbf{1}_U(\mathbf{R}\mathbf{1}_U)^{\sf T} + \mathbf{R}\mathbf{Z}\mathbf{R}^{\sf T}\,, \tag{18}$$

where $\mathbf{R} \in O(n)$ is an orthogonal matrix sampled according to the Haar measure. A direct calculation yields

$$\widehat{\mathbf{X}} = \beta\mathbf{v}\mathbf{v}^{\sf T} + \widetilde{\mathbf{Z}}, \tag{19}$$

where $\mathbf{v}$ is uniform on the n dimensional sphere, $\beta = L/\sqrt{n}$, and $\widetilde{\mathbf{Z}}$ is a GOE matrix (with off-diagonal entries of variance $1/n$). Furthermore, $\mathbf{v}$ and $\widetilde{\mathbf{Z}}$ are independent of one another.

Let $\mathbb{P}_{1,n}$ be the law of $\widehat{\mathbf{X}}$. Note that $\mathbb{P}_{1,n} = \mathbb{P}_\beta^{(k=2)}$ with $\beta = L/\sqrt{n}$. We can relate the detection problem of H_0 vs. $H_{1,L}$ to the detection problem of $\mathbb{P}_{0,n}$ vs. $\mathbb{P}_{1,n}$ as follows.

Lemma 4. *(a) If $\mathbb{P}_{1,n}$ is contiguous with respect to $\mathbb{P}_{0,n}$ then $H_{1,L}$ is spectrally contiguous with respect to H_0.*
(b) We have

$$\|Q_{0,n} - Q_{1,n}\|_{\rm TV} \leq \|\mathbb{P}_{0,n} - \mathbb{P}_{1,n}\|_{\rm TV}.$$

In view of Lemma 4, Theorem 1 is an immediate consequence of Theorem 3.

4 Proof of Theorem 3

The proof uses the following large deviations lemma, which follows, for instance, from [9, Proposition 2.3].

Lemma 5. *Let* $\mathbf{v}$ *a uniformly random vector on the unit sphere* $\mathbb{S}^{n-1}$ *and let* $\langle \mathbf{v}, \mathbf{e}_1\rangle$ *be its first coordinate. Then, for any interval* $[a,b]$ *with* $-1 \le a < b \le 1$

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\langle \mathbf{v}, \mathbf{e}_1\rangle \in [a,b]) = \max\Big\{\frac{1}{2}\log(1-q^2)\,:\, q\in[a,b]\Big\}\,. \tag{20}$$

Proof of Theorem 3. We denote by Λ the Radon-Nikodym derivative of $\mathbb{P}_\beta$ with respect to $\mathbb{P}_0$. By definition $\mathbb{E}_0\Lambda = 1$. It is easy to derive the following formula

$$\Lambda = \int \exp\Big\{-\frac{n\beta^2}{4} + \frac{n\beta}{2}\,\langle \mathbf{X}, \mathbf{v}^{\otimes k}\rangle\Big\}\,\mu_n(\mathrm{d}\mathbf{v})\,. \tag{21}$$

where μ_n is the uniform measure on $\mathbb{S}^{n-1}$. Squaring and using (11), we get

$$\begin{aligned}
\mathbb{E}_0\Lambda^2 &= e^{-n\beta^2/2}\int \mathbb{E}_0 \exp\Big\{\frac{n\beta}{2}\langle \mathbf{X}, {\mathbf{v}_1}^{\otimes k} + {\mathbf{v}_2}^{\otimes k}\rangle\Big\}\,\mu_n(\mathrm{d}\mathbf{v}_1)\mu_n(\mathrm{d}\mathbf{v}_2)\\
&= e^{-n\beta^2/2}\int \exp\Big\{\frac{n\beta^2}{4}\big\|{\mathbf{v}_1}^{\otimes k} + {\mathbf{v}_2}^{\otimes k}\big\|_F^2\Big\}\mu_n(\mathrm{d}\mathbf{v}_1)\mu_n(\mathrm{d}\mathbf{v}_2)\\
&= \int \exp\Big\{\frac{n\beta^2}{2}\langle \mathbf{v}_1, \mathbf{v}_2\rangle^k\Big\}\,\mu_n(\mathrm{d}\mathbf{v}_1)\mu_n(\mathrm{d}\mathbf{v}_2)\\
&= \int \exp\Big\{\frac{n\beta^2}{2}\langle \mathbf{v}, \mathbf{e}_1\rangle^k\Big\}\,\mu_n(\mathrm{d}\mathbf{v})\,,
\end{aligned} \tag{22}$$

where in the first step we used (11) and in the last step, we used rotational invariance.

Let $F_\beta : [-1,1] \to \overline{\mathbb{R}}$ be defined by

$$F_\beta(q) \equiv \frac{\beta^2 q^k}{2} + \frac{1}{2}\log(1-q^2)\,. \tag{23}$$

Using Lemma 5 and Varadhan's lemma, for any $-1 \le a < b \le 1$,

$$\int \exp\Big\{\frac{n\beta^2}{2}\langle \mathbf{v}, \mathbf{e}_1\rangle^k\Big\}\,\mathbb{I}(\langle \mathbf{v}, \mathbf{e}_1\rangle \in [a,b])\,\mu_n(\mathrm{d}\mathbf{v}) = \exp\Big\{n\max_{q\in[a,b]} F_\beta(q) + o(n)\Big\}\,. \tag{24}$$

It follows from the definition of $\beta_k^{\rm 2nd}$ that $\max_{|q|\ge\varepsilon} F_\beta(q) < 0$ for any $\varepsilon > 0$. Hence

$$\mathbb{E}_0\Lambda^2 \le \int \exp\Big\{\frac{n\beta^2}{2}\langle \mathbf{v}, \mathbf{e}_1\rangle^k\Big\}\,\mathbb{I}(|\langle \mathbf{v}, \mathbf{e}_1\rangle| \le \varepsilon)\,\mu_n(\mathrm{d}\mathbf{v}) + e^{-c(\varepsilon)n}\,, \tag{25}$$

for some $c(\varepsilon) > 0$ and all n large enough. Next notice that, under μ_n, $\langle \mathbf{v}, \mathbf{e}_1\rangle \overset{\rm d}{=} G/(G^2+Z_{n-1})^{1/2}$ where $G\sim \mathsf{N}(0,1)$ and Z_{n-1} is a χ^2 with $n-1$ degrees of freedom independent of G. Then, letting $Z_n \equiv G^2 + Z_{n-1}$ (a χ^2 with n degrees of freedom)

$$\begin{aligned}
\mathbb{E}_0\Lambda^2 &\le \mathbb{E}\Big\{\exp\Big(\frac{n\beta^2}{2}\frac{|G|^k}{Z_n^{k/2}}\Big)\,\mathbb{I}(|G/Z_n^{1/2}| \le \varepsilon)\Big\} + e^{-c(\varepsilon)n}\\
&\le \mathbb{E}\Big\{\exp\Big(\frac{n\beta^2}{2}\frac{|G|^k}{Z_n^{k/2}}\Big)\,\mathbb{I}(|G/Z_n^{1/2}| \le \varepsilon)\,\mathbb{I}(Z_{n-1} \ge n(1-\delta))\Big\}\\
&\quad + e^{n\beta^2\varepsilon^k/2}\mathbb{P}\big\{Z_{n-1} \le n(1-\delta)\big\} + e^{-c(\varepsilon)n}\\
&\le \mathbb{E}\Big\{\exp\Big(\frac{n^{1-(k/2)}\beta^2}{2(1-\delta)^{k/2}}\,|G|^k\Big)\mathbb{I}(|G|^2 \le 2\varepsilon n)\Big\} + e^{n\beta^2\varepsilon^k/2}\mathbb{P}\big\{Z_{n-1} \le n(1-\delta)\big\} + e^{-c(\varepsilon)n}\\
&= \frac{2}{\sqrt{2\pi}}\int_0^{2\varepsilon n} e^{C(\beta,\delta)n^{1-k/2}x^k - x^2/2}\mathrm{d}x + e^{n\beta^2\varepsilon^k/2}\mathbb{P}\big\{Z_{n-1} \le n(1-\delta)\big\} + e^{-c(\varepsilon)n}\,,
\end{aligned} \tag{26}$$

where $C(\beta,\delta) = \beta^2/(2(1-\delta)^{k/2})$. Now, for any $\delta > 0$, we can (and will) choose ε small enough so that both $e^{n\beta^2\varepsilon^k/2}\mathbb{P}\big\{Z_{n-1} \le n(1-\delta)\big\} \to 0$ exponentially fast (by tail bounds on χ^2 random variables) and, if $k \ge 3$, the argument of the exponent in the integral in the right hand side of (26) is bounded above by $-x^2/4$, which is possible since the argument vanishes at $x^* = 2C(\beta,\delta)n^{1/2}$. Hence, for any $\delta > 0$, and all n large enough, we have

$$\mathbb{E}_0\Lambda^2 \le \frac{2}{\sqrt{2\pi}} \int_0^{2\varepsilon n} e^{C(\beta,\delta)n^{1-k/2}x^k - x^2/2} \mathrm{d}x + e^{-c(\delta)n}\,, \tag{27}$$

for some $c(\delta) > 0$.

Now, for $k \ge 3$ the integrand in (27) is dominated by $e^{-x^2/4}$ and converges pointwise (as $n \to \infty$) to 1. Therefore, since $\mathbb{E}_0\Lambda^2 \ge (\mathbb{E}_0\Lambda)^2 = 1$,

$$k \ge 3: \qquad \lim_{n\to\infty} \mathbb{E}_0\Lambda^2 = 1\,. \tag{28}$$

For $k = 2$, the argument is independent of n and can be integrated immediately, yielding (after taking the limit $\delta \to 0$)

$$k = 2: \qquad \limsup_{n\to\infty} \mathbb{E}_0\Lambda^2 \le \frac{1}{\sqrt{1-\beta^2}}\,. \tag{29}$$

(Indeed, the above calculation implies that the limit exists and is given by the right-hand side.)

The proof is completed by invoking Lemma 2. □

5 Related work

In the classical $G(n, 1/2)$ planted clique problem, the computational problem is to find the planted clique (of cardinality k) in polynomial time, where we assume the location of the planted clique is hidden and is not part of the input. There are several algorithms that recover the planted clique in polynomial time when $k = C\sqrt{n}$ where $C > 0$ is a constant independent of n [2, 8, 10]. Despite significant effort, no polynomial time algorithm for this problem is known when $k = o(\sqrt{n})$. In the decision version of the planted clique problem, one seeks an efficient algorithm that distinguishes between a random graph distributed as $G(n, 1/2)$ or a random graph containing a planted clique of size $k \ge (2 + \delta)\log n$ (for $\delta > 0$; the natural threshold for the problem is the size of the largest clique in a random sample of $G(n, 1/2)$, which is asymptotic to $2\log n$ [14]). No polynomial time algorithm is known for this decision problem if $k = o(\sqrt{n})$.

As another example, consider the following setting introduced by [4] (see also [1]): one is given a realization of a n-dimensional Gaussian vector $\mathbf{x} := (\mathbf{x}_1, .., \mathbf{x}_n)$ with i.i.d. entries. The goal is to distinguish between the following two hypotheses. Under the first hypothesis, all entries in $\mathbf{x}$ are i.i.d. standard normals. Under the second hypothesis, one is given a family of subsets $C := \{S_1, ..., S_m\}$ such that for every $1 \le k \le m, S_k \subseteq \{1, ..., n\}$ and there exists an $i \in \{1, \dots, m\}$ such that, for any $\alpha \in S_i$, $\mathbf{x}_\alpha$ is a Gaussian random variable with mean $\mu > 0$ and unit variance whereas for every $\alpha \notin S_i$, $\mathbf{x}_\alpha$ is standard normal. (The second hypothesis does not specify the index i, only its existence). The main question is how large μ must be such that one can reliably distinguish between these two hypotheses. In [4], α are vertices in certain undirected graphs and the family C is a set of pre-specified paths in these graphs.

The Gaussian hidden clique problem is related to various applications in statistics and computational biology [6, 18]. That detection is statistically possible when $L \gg \log n$ was established in [1]. In terms of *polynomial time* detection, [8] show that detection is possible when $L = \Theta(\sqrt{n})$ for the symmetric cases. As noted, no polynomial time algorithm is known for the Gaussian hidden clique problem when $k = o(\sqrt{n})$. In [1, 20] it was hypothesized that the Gaussian hidden clique problem should be difficult when $L \ll \sqrt{n}$.

The closest results to ours are the ones of [22]. In the language of the present paper, these authors consider a rectangular matrix of the form $\mathbf{X} = \lambda\, \mathbf{v}_1\mathbf{v}_2^{\mathsf{T}} + \mathbf{Z} \in \mathbb{R}^{n_1\times n_2}$ whereby $\mathbf{Z}$ has i.i.d. entries $\mathbf{Z}_{ij} \sim \mathsf{N}(0, 1/n_1)$, $\mathbf{v}_1$ is deterministic of unit norm, and $\mathbf{v}_2$ has entries which are i.i.d. $\mathsf{N}(0, 1/n_1)$, independent of $\mathbf{Z}$. They consider the problem of testing this distribution against $\lambda = 0$. Setting $c = \lim_{n\to\infty} \frac{n_1}{n_2}$, it is proved in [22] that the distribution of the singular values of $\mathbf{X}$ under the null and the alternative are mutually contiguous if $\lambda < \sqrt{c}$ and not mutually contiguous if $\lambda > \sqrt{c}$. While [22] derive some more refined results, their proofs rely on advanced tools from random matrix theory [13], while our proof is simpler, and generalizable to other settings (e.g. tensors).

References

[1] L. Addario-Berry, N. Broutin, L. Devroye, G. Lugosi. On combinatorial testing problems. *Annals of Statistics* 38(5) (2011), 3063–3092.

[2] N. Alon, M. Krivelevich and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures and Algorithms* 13 (1998), 457–466.

[3] G. W. Anderson, A. Guionnet and O. Zeitouni. An introduction to random matrices. Cambridge University Press (2010).

[4] E. Arias-Castro, E. J., Candès, H. Helgason and O. Zeitouni. Searching for a trail of evidence in a maze. *Annals of Statistics* 36 (2008), 1726–1757.

[5] A. Auffinger, G. Ben Arous, and J. Cerny. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics* 66(2) (2013), 165–201.

[6] S. Balakrishnan, M. Kolar, A. Rinaldo, A. Singh, and L. Wasserman. Statistical and computational tradeoffs in biclustering. *NIPS* Workshop on Computational Trade-offs in Statistical Learning (2011).

[7] S. Bhamidi, P.S. Dey, and A.B. Nobel. Energy landscape for large average submatrix detection problems in Gaussian random matrices. arXiv:1211.2284.

[8] Y. Deshpande and A. Montanari. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Foundations of Computational Mathematics* (2014), 1–60

[9] A. Dembo and O. Zeitouni. Matrix optimization under random external fields. arXiv:1409.4606

[10] U. Feige and R. Krauthgamer. Finding and certifying a large hidden clique in a semi-random graph. *Random Struct. Algorithms* 162(2) (1999), 195–208.

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

[12] Z. Füredi and J. Komlós, The eigenvalues of random symmetric matrices. *Combinatorica* 1 (1981), 233–241.

[13] A. Guionnet and M. Maida. A Fourier view on R-transform and related asymptotics of spherical integrals. *Journal of Functional Analysis* 222 (2005), 435–490.

[14] G. R. Grimmett and C. J. H. McDiarmid. On colouring random graphs. *Math. proc. Cambridge Philos. Soc.* 77 (1975), 313–324.

[15] D. Hsu, S. M. Kakade, and T. Zhang. A spectral algorithm for learning hidden Markov models. *Journal of Computer and System Sciences* 78.5 (2012): 1460-1480.

[16] M. Jerrum. Large cliques elude the Metropolis process. *Random Struct. Algorithms* 3(4) (1992), 347–360.

[17] A. Knowles and J. Yin, The isotropic semicircle law and deformation of Wigner matrices. *Communications on Pure and Applied Mathematics* 66(11) (2013), 1663–1749.

[18] M. Kolar, S. Balakrishnan, A. Rinaldo, and A. Singh. Minimax localization of structural information in large noisy matrices. *Neural Information Processing Systems (NIPS)*, (2011), 909–917.

[19] M. Talagrand. Free energy of the spherical mean field model. *Probability theory and related fields* 134(3) (2006), 339–382.

[20] Z Ma and Y Wu. Computational barriers in minimax submatrix detection. arXiv:1309.5914.

[21] A. Montanari and E. Richard. A Statistical Model for Tensor PCA. *Neural Information Processing Systems (NIPS)* (2014), 2897–2905.

[22] A. Onatski, M. J. Moreira, M. Hallin, et al. Asymptotic power of sphericity tests for high-dimensional data. *The Annals of Statistics* 41(3) (2013), 1204–1231.

[23] W. C. Waterhouse. The absolute-value estimate for symmetric multilinear forms. *Linear Algebra and its Applications* 128 (1990), 97–105.