Random Matrix Contiguity

Alex Wein

May 11, 2016

1 Introduction

In the "spike model" for random matrices, it is well known that the top eigenvalue of a GOE matrix is affected by a planted 'spike' if and only if the size of the spike exceeds a certain threshold. This leaves open the question of what happens when the spike is below the threshold – is there another way (other than the top eigenvalue) to detect the spike (e.g. by looking at other eigenvalues or gaps between eigenvalues)? In this report I will give a simple proof that it is in fact statistically impossible to detect the spike below the threshold. This is based on joint research with Will Perry.

2 Spike Model

A "spiked Wigner matrix" takes the form $Y = \lambda x x^\top + \frac{1}{\sqrt{n}} W$ where x is a unit vector in $\mathbb{R}^n$ and W is an $n \times n$ GOE matrix (normalized such that the off-diagonals have variance 1). Here λ is a parameter that controls the size of the spike. It is known that the top eigenvalue undergoes a phase transition at the critical value $\lambda = 1$, namely:

Theorem 2.1 ([FP06]). *Let $Y = \lambda x x^\top + \frac{1}{\sqrt{n}} W$ as above.*

- *If $\lambda \le 1$ then $\lambda_{\max}(Y) \to 2$ as $n \to \infty$,*
- *and if $\lambda > 1$ then $\lambda_{\max}(Y) \to \lambda + \frac{1}{\lambda} > 2$ as $n \to \infty$.*

Recall that $\lambda_{\max}(\frac{1}{\sqrt{n}} W) \to 2$, so this means that the spike affects the top eigenvalue if and only if $\lambda > 1$. Our main result will show that when $\lambda < 1$, it is statistically impossible to detect the spike. (For simplicity we do not consider the boundary case $\lambda = 1$.)

3 Contiguity

The proof of our main result will rely on the notion of contiguity (see [Jan95]), which was introduced by Le Cam as the asymptotic analogue of absolute continuity. We consider two sequences of probability measures $\{\mathbb{P}_n\}$ and $\{\mathbb{Q}_n\}$ such that for each n, $\mathbb{P}_n$ and $\mathbb{Q}_n$ are defined on the same probability space. In our case, we will be interested in the following two distributions over $n \times n$ matrices Y_n:

- $\mathbb{P}_n$: $Y_n = \lambda x x^\top + \frac{1}{\sqrt{n}} W$ where $x \sim \mathcal{D}_n$
- $\mathbb{Q}_n$: $Y_n = \frac{1}{\sqrt{n}} W$.

In other words: under $\mathbb{P}_n$, Y_n is a spiked Wigner matrix where the spike is drawn from some prior $\mathcal{D}_n$; and under $\mathbb{Q}_n$, Y_n is just an (un-spiked) Wigner matrix. We will take $\mathcal{D}_n$ to be the uniform prior over unit vectors in $\mathbb{R}^n$, but our techniques also extend to other priors (e.g. vectors with entries $\{\pm 1\}$). Now we are ready to define contiguity.

Definition 3.1. We say $\{\mathbb{P}_n\}$ is *contiguous* to $\{\mathbb{Q}_n\}$ if whenever $\mathbb{Q}_n(A_n) \to 0$ as $n \to \infty$ for a sequence of events $\{A_n\}$, we also have $\mathbb{P}_n(A_n) \to 0$. We denote this by $\mathbb{P}_n \triangleleft \mathbb{Q}_n$.

Our main result shows contiguity for the specific $\{\mathbb{P}_n\}$, $\{\mathbb{Q}_n\}$ defined above.

Theorem 3.2 (main)**.** *For $\{\mathbb{P}_n\}$ and $\{\mathbb{Q}_n\}$ defined above (spiked Wigner matrices), if $\lambda < 1$ then we have $\mathbb{P}_n \triangleleft \mathbb{Q}_n$.*

The reason we are interested in contiguity is because of its implications for non-distinguishability of the two distributions in the following sense. Suppose we generate a random value Y_n as follows: with probability $\frac{1}{2}$, Y_n is sampled from $\mathbb{P}_n$, and with probability $\frac{1}{2}$, Y_n is sampled from $\mathbb{Q}_n$. Also suppose we have a "distinguisher" $\mathcal{A}_n$ that takes Y_n and tries to guess which of the two distributions it came from. An immediate consequence of contiguity is that if $\mathbb{P}_n \triangleleft \mathbb{Q}_n$ then there is no distinguisher $\mathcal{A}_n$ that guesses correctly with probability $1 - o(1)$ as $n \to \infty$. To prove this, consider the event A_n that $\mathcal{A}_n$ guesses "$\mathbb{P}_n$". If the distinguisher succeeds with probability $1 - o(1)$ then we must have $\mathbb{Q}_n(A_n) \to 0$ (i.e. if Y_n comes from $\mathbb{Q}_n$ then $\mathcal{A}_n$ should not guess "$\mathbb{P}_n$"). But by contiguity this implies $\mathbb{P}_n(A_n) \to 0$, which is a contradiction. This gives the following corollary.

Corollary 3.3. *For $\{\mathbb{P}_n\}$ and $\{\mathbb{Q}_n\}$ defined above (spiked Wigner matrices), if $\lambda < 1$ then no distinguisher succeeds with probability $1 - o(1)$ as $n \to \infty$.*

In other words, there is no test that can reliably detect the presence of the spike (when $\lambda < 1$). Note that this only rules out distinguishers that succeed with high probability $1 - o(1)$. There could still be a distinguisher that succeeds with, say, some constant probability larger than $\frac{1}{2}$.

4 Proof of Main Result

In this section we prove our main result (Theorem 3.2): if $\lambda < 1$ then $\mathbb{P}_n \triangleleft \mathbb{Q}_n$. A related result can be found in [KXZ16]. One advantage of our proof is that it is very simple.

We start with a crucial lemma that gives us a concrete way to prove contiguity: if we can show that a particular second moment is finite, then contiguity follows.

Lemma 4.1. *Let $\{\mathbb{P}_n\}$ and $\{\mathbb{Q}_n\}$ be two sequences of probability measures. If the second moment*

$$\mathbb{E}_{\mathbb{Q}_n}\left[\left(\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\right)^2\right]$$

remains bounded as $n \to \infty$ then $\mathbb{P}_n \triangleleft \mathbb{Q}_n$.

In our case, where $\mathbb{P}_n$ and $\mathbb{Q}_n$ are continuous distributions supported everywhere, say with densities p_n and q_n, the second moment is just $\mathbb{E}_{Y \sim \mathbb{Q}_n}\left[\left(\frac{p_n(Y)}{q_n(Y)}\right)^2\right]$.

Proof. Let $\{A_n\}$ be a sequence of events such that $\mathbb{Q}_n(A_n) \to 0$ as $n \to \infty$. Using Cauchy-Schwarz,

$$\begin{aligned}\mathbb{P}_n(A_n) = \int_{A_n} d\mathbb{P}_n = \int_{A_n} \left(\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\right) d\mathbb{Q}_n &\le \sqrt{\int_{A_n} \left(\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\right)^2 d\mathbb{Q}} \cdot \sqrt{\int_{A_n} d\mathbb{Q}_n} \\ &\le \sqrt{\mathbb{E}_{\mathbb{Q}_n}\left(\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\right)^2} \cdot \sqrt{\mathbb{Q}_n(A_n)}.\end{aligned}$$

The first factor is bounded and the second goes to 0, so $\mathbb{P}_n(A_n) \to 0$. ☐

4.1 Proof of Theorem 3.2

We can now prove our main theorem by computing the second moment $\mathbb{E}_{\mathbb{Q}_n}\left[\left(\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\right)^2\right]$ for our particular choice of $\mathbb{P}_n$ and $\mathbb{Q}_n$ (spiked Wigner matrices).

$$\begin{aligned}
\frac{d\mathbb{P}_n}{d\mathbb{Q}_n} &= \frac{\mathbb{E}_{x\sim\mathcal{D}_n} \exp\left(-\frac{n}{2}\sum_{i<j}(Y_{ij}-\lambda x_i x_j)^2 - \frac{n}{4}\sum_i (Y_{ii}-\lambda x_i^2)^2\right)}{\exp\left(-\frac{n}{2}\sum_{i<j} Y_{ij}^2 - \frac{n}{4}\sum_i Y_{ii}^2\right)} \\
&= \mathbb{E}_{x\sim\mathcal{D}_n} \exp\left(n\sum_{i<j}\lambda Y_{ij}x_i x_j - \frac{n}{2}\sum_{i<j}\lambda^2 x_i^2 x_j^2 + \frac{n}{2}\sum_i \lambda Y_{ii}x_i^2 - \frac{n}{4}\sum_i \lambda^2 x_i^4\right) \\
&= \mathbb{E}_{x\sim\mathcal{D}_n} \exp\left(n\sum_{i<j}\lambda Y_{ij}x_i x_j + \frac{n}{2}\sum_i \lambda Y_{ii}x_i^2 - \frac{n}{4}\sum_{i,j}\lambda^2 x_i^2 x_j^2\right) \\
&= \mathbb{E}_{x\sim\mathcal{D}_n} \exp\left(n\sum_{i<j}\lambda Y_{ij}x_i x_j + \frac{n}{2}\sum_i \lambda Y_{ii}x_i^2 - \frac{n}{4}\lambda^2\right) \quad \text{since } \|x\|=1
\end{aligned}$$

$$\left(\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\right)^2 = \exp\left(-\frac{n\lambda^2}{2}\right)\mathbb{E}_{x,x'\sim\mathcal{D}_n} \exp\left(n\sum_{i<j}\lambda Y_{ij}x_i x_j + \frac{n}{2}\sum_i \lambda Y_{ii}x_i^2 + n\sum_{i<j}\lambda Y_{ij}x'_i x'_j + \frac{n}{2}\sum_i \lambda Y_{ii}(x'_i)^2\right)$$

where x, x' are drawn independently from $\mathcal{D}_n$.

$$\mathbb{E}_{Y\sim\mathbb{Q}_n}\left(\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}\right)^2 = \exp\left(-\frac{n\lambda^2}{2}\right)\mathbb{E}_{x,x'\sim\mathcal{D}_n}\mathbb{E}_{Y\sim\mathbb{Q}_n} \exp\left(n\sum_{i<j}\lambda Y_{ij}(x_i x_j + x'_i x'_j) + \frac{n}{2}\sum_i \lambda Y_{ii}(x_i^2 + (x'_i)^2)\right)$$

use the Gaussian moment-generating function:

$$\begin{aligned}
&= \exp\left(-\frac{n\lambda^2}{2}\right)\mathbb{E}_{x,x'\sim\mathcal{D}_n} \exp\left(\sum_{i<j}\frac{n\lambda^2}{2}(x_i x_j + x'_i x'_j)^2 + \sum_i \frac{n\lambda^2}{4}(x_i^2 + (x'_i)^2)^2\right) \\
&= \exp\left(-\frac{n\lambda^2}{2}\right)\mathbb{E}_{x,x'\sim\mathcal{D}_n} \exp\left(\frac{n\lambda^2}{4}\sum_{i,j}(x_i x_j + x'_i x'_j)^2\right) \\
&= \exp\left(-\frac{n\lambda^2}{2}\right)\mathbb{E}_{x,x'\sim\mathcal{D}_n} \exp\left(\frac{n\lambda^2}{4}\sum_{i,j}(x_i^2 x_j^2 + 2x_i x'_i x_j x'_j + (x'_i)^2(x'_j)^2)\right) \\
&= \exp\left(-\frac{n\lambda^2}{2}\right)\mathbb{E}_{x,x'\sim\mathcal{D}_n} \exp\left(\frac{n\lambda^2}{2}(1 + \langle x, x'\rangle^2)\right) \quad \text{since } \|x\|=1 \\
&= \mathbb{E}_{x,x'\sim\mathcal{D}_n} \exp\left(\frac{n\lambda^2}{2}\langle x, x'\rangle^2\right) \\
&= \mathbb{E}_{x\sim\mathcal{D}_n} \exp\left(\frac{n\lambda^2}{2}\langle x, e_1\rangle^2\right) \quad \text{by symmetry (here } e_1 = (1,0,0,\ldots))
\end{aligned}$$

For large n, the distribution of $\langle x, e_1 \rangle$ approaches $\mathcal{N}(0, 1/n)$ and so the distribution of $\langle x, e_1 \rangle^2$ approaches $\frac{1}{n}\chi_1^2$. Using the chi-squared moment-generating function:

$$\begin{aligned} &= \left(1 - 2 \cdot \frac{\lambda^2}{2}\right)^{-1/2} \\ &= \frac{1}{\sqrt{1 - \lambda^2}} \end{aligned}$$

which is bounded provided $\lambda < 1$. Our main result now follows from Lemma 4.

References

[FP06] D. Féral and S. Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in Mathematical Physics*, 272(1):185–228, 2006.

[Jan95] S. Janson. Random regular graphs: asymptotic distributions and contiguity. *Combinatorics, Probability and Computing*, 4(04):369–405, 1995.

[KXZ16] F. Krzakala, J. Xu, and L. Zdeborová. Mutual Information in Rank-One Matrix Estimation. *arXiv:1603.08447*, March 2016.