AM/ES 254: Mathematics of High-Dimensional Information Processing and Learning
Fall 2022
Professor: Yue M. Lu

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

Homework 2

Due: Wednesday Oct. 19, 2022 at 1 pm

Exercise 1. COMPUTER EXERCISE: THE CIRCULAR LAW AND ITS UNIVERSALITY

In our lecture we considered a symmetric random matrix H that had independent entries in its upper right part. The symmetry of H makes sure that its eigenvalues are all real-valued. If we allow H to be asymmetric, e.g., having fully independent entries, it will then have complex eigenvalues. In this case, the densities of the eigenvalues of H, with appropriate scaling, will approximate a uniform distribution within the unit circle. This is the so-called *circular law*.
In this computer exercise, write a program to illustrate the circular law and demonstrate the universality phenomenon by using the following constructions of the random matrix H:

1. $H \in \mathbb{R}^{n\times n}$ has i.i.d.Gaussian entries $\mathcal{N}(0, 1/n)$.
2. $\{H_{ij}\}$ are i.i.d. samples of a discrete random variable that takes values from $\left\{\pm\frac{1}{\sqrt{n}}\right\}$ with equal probability.
3. Fill an $n \times n$ matrix A, in a column-wise fashion, with the first n^2 digits of $\pi = 3.1415926\ldots$. Let $H = a(A - b11^\top)$, where 1 denotes the all-one vector, and a, b are two scalars chosen such that the empirical mean of the entries of H is zero and the empirical variance of the entries of H is equal to $1/n$. Now repeat the above experiment by replacing π with $\sqrt{2}$ and e, respectively.

You can start by working with $n = 500$ in your simulations, but you will get better-lookingx images if you work with larger values of n (e.g. $n = 2000$).

Exercise 2. CONCENTRATION OF RANDOM MATRICES VIA ϵ-NETS (DIFFICULTY LEVEL: ☕☕)

In this problem, we will explore a technique often used to control the concentration of random matrices in the spectral norm.

Consider the following symmetric random matrix:

$$\Psi \overset{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot a_i a_i^T$$

where we assume that

- the vectors $a_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_n)$;
- the scalars σ_i are *i.i.d.* random variables with bounded support ($\mathbb{P}[|\sigma| \le B] = 1$ for some finite constant $B > 0$);
- the vectors a_i and the scalars σ_i are mutually independent.

The objective of this exercise will be to estimate, with high probability, the deviation of Ψ from its expectation in the spectral norm:

$$\|\Psi - \mathbb{E}[\Psi]\|_2 \le \delta'$$

for some small positive δ'.

We start recalling the following fact on the spectral norm of a symmetric matrix.

Lemma 1 (Spectral norm of a Symmetric Matrix)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then*

$$\|A\|_2 = \max_{v\in\mathbb{S}^{n-1}} |v^\top A v|$$

where $\mathbb{S}^{n-1} = \{v \in \mathbb{R}^n \ : \ \|v\|_2 = 1\}$ is the hypersphere in $\mathbb{R}^n$.

1. Show that there exist $c, k > 0$ such that $\forall v \in \mathbb{S}^{n-1}$ we have

$$\mathbb{P}\Big(|v^\top(\Psi - \mathbb{E}\Psi)v| \le t\Big) \ge 1 - 2\exp\Big(-c\cdot\min\Big\{\frac{t}{k}, \frac{t^2}{k^2}\Big\}\cdot m\Big) \tag{1}$$

We would like now to extend the above bound to the whole hypersphere and from

$$\mathbb{P}\Big(\max_{v\in\mathbb{S}^{n-1}} |v^\top(\Psi - \mathbb{E}\Psi)v| \le t\Big) = \mathbb{P}\Big(|v^\top(\Psi - \mathbb{E}\Psi)v| \le t;\ \forall v \in \mathbb{S}^{n-1}\Big).$$

obtain our sought bound. Notice though that (1) is a bound for every *fixed* $v \in \mathbb{S}^{n-1}$, while we require a bound *uniform* in $v \in \mathbb{S}^{n-1}$ (in the first case $\forall v$ is outside the probability bound, in the second case it is inside!). For this, we are going to use an ϵ-net argument.

Definition 1 (ϵ-net)**.** *Let $\epsilon > 0$. Then $V_\epsilon \subset \mathbb{S}^{n-1}$ is called an ϵ-net, if for every $v \in \mathbb{S}^{n-1}$ there exists $v_0 \in V$ such that $\|v - v_0\|_2 \le \epsilon$.*

Lemma 2. *Let $\epsilon > 0$. Then there exists an ϵ-net $V_\epsilon \subset \mathbb{S}^{n-1}$ such that V_ϵ is finite (contains a finite number of vectors) and $|V_\epsilon| \le (\frac{2}{\epsilon} + 1)^n$.*

In the next steps, we are going to "discretize" the hypersphere $\mathbb{S}^{n-1}$ using an ϵ-net and show that the maximum over $\mathbb{S}^{n-1}$ can be controlled using the bound over the ϵ-net.

2. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and V_ϵ an ϵ-net. Show that

$$\max_{v\in\mathbb{S}^{n-1}} |v^\top A v| \le \frac{1}{1-2\epsilon} \max_{v\in V_\epsilon} |v^\top A v|.$$

3. Let $V_{\frac{1}{4}}$ a 1/4-net. Argue that

$$\mathbb{P}\Big(\max_{v\in V_\epsilon} |v^\top(\Psi - \mathbb{E}\Psi)v| \ge \frac{t}{2}\Big) \le 2\cdot 9^n \cdot \exp\Big(-c\cdot\min\Big\{\frac{t}{k}, \frac{t^2}{k^2}\Big\}\cdot m\Big)$$

4. Let $\delta \in (0, 1]$. Show that for some constants $c, C > 0$, if $m \ge C\cdot n/\delta^2$ then

$$\mathbb{P}\Big(\|\Psi - \mathbb{E}\Psi\|_2 \ge c\cdot k\cdot\delta\Big) \le 2e^{-Cn}.$$

   Notice that the above says that for vectors a_i in $\mathbb{R}^n$, we need about $m \gtrsim n$ number of samples of a_i and σ_i to estimate the expectation of Ψ.

Exercise 3. Fluctuations of the Largest Eigenvalue (Difficulty level: ☕☕☕☕)

Let $Y \in \mathbb{R}^{n\times n}$ be a noisy realization of the rank-1 matrix $e_1 e_1^\top$ generated as follows:

$$Y = \theta\cdot e_1 e_1^\top + W, \tag{2}$$

where $\theta \geq 0$ is the signal-to-noise ratio parameter and W is a symmetric noise matrix whose upper triangular entries are sampled independently from the distribution:

$$W_{ij} \sim \begin{cases} \mathcal{N}(0, 1/n) & : i < j \\ \mathcal{N}(0, 2/n) & : i = j \end{cases}.$$

Let $\lambda_1^{(n)}$ denote the largest eigenvalue of Y. Recall in the lecture, we showed that:

$$\lambda_1^{(n)} \to \Lambda(\theta) \stackrel{\text{def}}{=} \begin{cases} 2 & : \theta \leq 1 \\ \theta + \frac{1}{\theta} & : \theta > 1 \end{cases}.$$

In this problem, we will analyze the fluctuations of $\lambda_1^{(n)}$ when $\underline{\theta > 1}$. We will show that:

$$\sqrt{n}(\lambda_1^{(n)} - \Lambda(\theta)) \to \mathcal{N}\left(0, \sigma^2(\theta)\right),$$

and derive a formula for $\sigma^2(\theta)$. The goal of this exercise is to practice heuristic derivations of such results, so you only need to provide intuitive/heuristic arguments for the following problems; rigorous proofs are not required.

1. Using a first-order Taylor's expansion, derive the approximation:

$$\sqrt{n} \cdot (\lambda_1^{(n)} - \Lambda(\theta)) \approx \sqrt{n} \cdot \frac{F(\Lambda(\theta); \theta) - F_n(\Lambda(\theta); \theta)}{\partial_\lambda F_n(\Lambda(\theta); \theta)},$$

where:

$$F_n(\lambda; \theta) \stackrel{\text{def}}{=} \lambda - \theta - W_{11} - g^\top (\lambda I_{n-1} - W_{\setminus 1})^{-1} g,$$

$$F(\lambda; \theta) \stackrel{\text{def}}{=} \lambda - \theta - \int_{\mathbb{R}} \frac{\mu_{sc}(\mathrm{d}x)}{\lambda - x} = \lambda - \theta - \frac{\lambda - \sqrt{\lambda^2 - 4}}{2}.$$

In the above display, $g \stackrel{\text{def}}{=} (W_{21}, \dots, W_{n1})^\top \in \mathbb{R}^{n-1}$ and $W_{\setminus 1}$ is the principal submatrix of W obtained by deleting the first row and column.

Hint: Recall in the lecture notes we showed that when $\theta > 1$, $\lambda_1^{(n)}$ is the solution of the equation $F_n(\lambda_1^{(n)}; \theta) = 0$. Write down a first-order Taylor expansion of this equation around $\Lambda(\theta)$, the limiting value of $\lambda_1^{(n)}$.

2. Provide a heuristic derivation of the Gaussian approximation:

$$\sqrt{n} \cdot (F(\Lambda(\theta); \theta) - F_n(\Lambda(\theta); \theta)) \to \mathcal{N}\left(0, s^2(\theta)\right),$$

where:

$$s^2(\theta) = 2 + 2 \int_{\mathbb{R}} \frac{\mu_{sc}(\mathrm{d}x)}{(\Lambda(\theta) - x)^2}. \tag{3}$$

You may use the following fact without justification in your derivation.

Fact 1. *For any $\lambda > 2$ (independent of n),*

$$\frac{\mathrm{Tr}[(\lambda I_n - W)^{-1}]}{n} - \int_{\mathbb{R}} \frac{\mu_{sc}(\mathrm{d}x)}{\lambda - x} = O_\prec(n^{-1}).$$

Hint. An important term in the expression for $\sqrt{n} \cdot (F(\Lambda(\theta); \theta) - F_n(\Lambda(\theta); \theta))$ is $g^\top (\lambda I_{n-1} - W_{\setminus 1})^{-1} g$. Use the rotation trick to show that:

$$g^\top (\lambda I_{n-1} - W_{\setminus 1})^{-1} g \stackrel{\text{d}}{=} \frac{1}{n} \sum_{i=1}^{n-1} \frac{z_i^2}{\lambda - \gamma_i},$$

where $z_{1:n-1} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ and $\gamma_1, \dots, \gamma_{n-1}$ are the eigenvalues of $W_{\setminus 1}$. Apply the Central Limit Theorem to the random variable on the RHS (after suitable centering and rescaling).

3. Provide a closed-form formula for $s^2(\theta)$ in terms of θ by computing the integral in (3).

   Hint. Recall that from the lecture notes we know that,

$$\int_{\mathbb{R}} \frac{\mu_{sc}(\mathrm{d}x)}{\lambda - x} = \frac{\lambda - \sqrt{\lambda^2 - 4}}{2}. \tag{4}$$

   Think about how this can be used to derive a formula for:

$$\int_{\mathbb{R}} \frac{\mu_{sc}(\mathrm{d}x)}{(\lambda - x)^2}.$$

4. Provide a heuristic argument to show that:

$$\partial_\lambda F_n(\Lambda(\theta); \theta) \to \kappa(\theta),$$

   for some constant $\kappa(\theta)$. Provide a formula for $\kappa(\theta)$ in terms of θ.

5. Combine your results from parts (1-4) to justify that $\sqrt{n}(\lambda_1^{(n)} - \Lambda(\theta)) \to \mathcal{N}\left(0, \sigma^2(\theta)\right)$ and provide a formula for $\sigma^2(\theta)$.

6. Check the validity of your result using numerical simulations: set $\theta = 4, n = 1000$ and sample $m = 1000$ i.i.d. random matrices $Y[1], \ldots, Y[m]$ from the model in (2). Compute the largest eigenvalues $\lambda_1^{(n)}[1], \ldots, \lambda_1^{(n)}[m]$ of these matrices. Plot a histogram of:

$$\sqrt{n} \cdot \frac{\lambda_1^{(n)}[1] - \Lambda(\theta)}{\sigma(\theta)}, \ldots, \sqrt{n} \cdot \frac{\lambda_1^{(n)}[m] - \Lambda(\theta)}{\sigma(\theta)},$$

   and compare it with the density function of $\mathcal{N}(0, 1)$.