Random Matrix Theory I

1 Top Eigenvalue Problem

Consider a matrix $J \in \mathbb{R}^{n\times n}$ with $J_{ij} = J_{ji}$, $J_{ij} \sim \mathcal{N}(0, \frac{1}{n})$ and $J_{ii} \sim \mathcal{N}(0, \frac{2}{n})$. J is called a **GOE matrix**. In these next two lectures, we aim to use the Replica method to compute $\lambda_{\max}(J)$, which we can formulate as an optimization problem via:

$$\max_{\|x\|_2=1} \langle x, Jx\rangle$$

We will use slightly informal arguments for some calculations, without providing additional details, since we are anyway not doing a fully rigorous proof in the sense of math.

1.1 Warmup: Surface Area of High-Dimensional Spheres

As a warmup, we compute $\log \mathrm{SA}(\sqrt{n}S^{n-1})$ to leading order. This calculation will illustrate an informal idea related to "equivalence of ensembles" in statistical mechanics. We recall some facts:

1. Gaussian integrals:
$$\int_{\mathbb{R}} e^{-x^2/2} = \sqrt{2\pi} \qquad\qquad \int_{\mathbb{R}^n} e^{-\|x\|_2^2/2} = (2\pi)^{n/2}$$
2. Poincaré's inequality:
$$\mathrm{Var}(\|x\|_2) \le \mathbb{E}\left|\nabla\|x\|_2\right|^2 = \mathbb{E}\left|\frac{x}{\|x\|_2}\right|^2 = 1 \ll \sqrt{n}$$
3. If $x \sim \mathcal{N}(0, I_n)$, $\|x\|_2^2 \sim \chi^2(n)$, so the typical size of $\|x\|_2$ is $\sqrt{n}$
4. "Equivalence of Ensembles": Say $x \sim \mathcal{N}(0, I_n)$. We rewrite it as:
$$x = \frac{x}{\|x\|_2} \cdot \|x\|_2$$
$\frac{x}{\|x\|_2}$ is a unit-norm direction vector on S^{n-1}, and $\|x\|_2 \approx \sqrt{n}$. So therefore:
$$x \sim \mathcal{N}(0, I_n) \text{ is approximated as } \mathrm{Unif}(\sqrt{n}S^{n-1})$$
and $\mathcal{N}(0, I_n) \approx \mathrm{Unif}(\sqrt{n}S^{n-1})$.

Remark 1. The Gaussian pdf $\propto e^{-\|x\|_2^2/2}$ can be interpreted as $e^{-\beta H(x)}$ with $\beta = 1$ and $H(x) = \frac{\|x\|_2^2}{2}$. Informally, the inverse temperature β plays the role of a "Lagrange multiplier" which enforces $H(x) \approx n/2$. See a textbook for more explanation.

With these facts:

$$\log \mathrm{SA}(\sqrt{n}S^{n-1}) = \log \int_{\|x\|_2=\sqrt{n}} 1\, dx \approx \log \int_{\|x\|_2\approx\sqrt{n}} e^{-\|x\|_2^2/2} e^{n/2}\, dx$$

where we approximated $1 \approx e^{-\|x\|^2}/2e^{n/2}$ because $\|x\|_2 \approx \sqrt{n}$. Since $\|x\|_2$ is concentrated about $\sqrt{n}$,

$$\frac{\log \int_{\|x\|_2 \approx \sqrt{n}} e^{-\|x\|_2^2/2} e^{n/2}\, dx}{\log \int_{\mathbb{R}^n} e^{-\|x\|_2^2/2} e^{n/2}\, dx} \approx 1$$

So:

$$\log \mathrm{SA}(\sqrt{n} S^{n-1}) \approx \log \int_{\mathbb{R}^n} e^{-\|x\|_2^2/2} e^{n/2}\, dx = \log(2\pi e)^{n/2} = \frac{n}{2}\log(2\pi e)$$

So therefore:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{SA}(\sqrt{n} S^{n-1}) = \frac{1}{2}\log(2\pi e)$$

1.2 Heuristics of $\lambda_{\max}(J)$ via Replica Calculation

$$\mathbb{E}\left[\frac{1}{2}\lambda_{\max}(E)\right] = \mathbb{E}\left[\max_{\|x\|_2=1} \frac{\langle x, Jx\rangle}{2}\right] \overset{(*)}{=} \lim_{\beta\to 0} \mathbb{E}\left[\frac{1}{\beta n} \log \int_{\mathrm{Unif}(\sqrt{n}S^{n-1})} e^{\beta\langle x, Jx\rangle/2}\right]$$

The equality in $(*)$ is non-trivial to see, and might show up on the homework. We claim that the integral in the expectation concentrates. Via Poincaré's inequality,

$$\mathrm{Var}\left(\log \int e^{\beta\langle x, Jx\rangle/2}\right) \leq \mathbb{E}\left[\left|\nabla_J\left(\max_{\|x\|_2=1} \frac{\langle x, Jx\rangle}{2}\right)\right|^2\right]$$

By direct calculation,

$$\nabla_J\left(\max_{x\in\sqrt{n}S^{n-1}} \frac{\langle x, Jx\rangle}{2}\right) = \nabla_J\left(\max_{x\in\sqrt{n}S^{n-1}} \frac{\langle xx^T, J\rangle}{2}\right) = \frac{xx^T}{2} \sim \sqrt{n}$$

Thus,

$$\mathrm{Var}\left(\frac{1}{\beta n}\log \int e^{\beta\langle x, Jx\rangle/2}\right) = \frac{1}{\beta^2 n^2}\mathrm{Var}\left(\log \int e^{\beta\langle x, Jx\rangle/2}\right) \leq \frac{1}{\beta^2 n^2}\cdot n = \frac{1}{\beta^2 n}$$

So the integral concentrates. Now for the Replica trick:

1. $\mathbb{E}[\log Z] = \lim_{k\to 0} \frac{\log \mathbb{E}[Z^k]}{k}$

2. Take a high-dimensional limit:

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\log Z] = \lim_{n\to\infty}\lim_{k\to\infty} \frac{1}{nk}\log \mathbb{E}[Z^k] = \lim_{k\to\infty}\lim_{n\to\infty} \frac{1}{nk}\log \mathbb{E}[Z^k]$$

   The second equality is a guess (i.e., not fully justified).

3. Now, we guess the formula for $\lim_{n\to\infty} \frac{1}{nk}\log \mathbb{E}[Z^k]$ for $k \approx 0$ from a formula for $k \in \mathbb{Z}$. Let μ be the uniform measure on $\sqrt{n}S^{n-1}$. Then:

$$\begin{aligned}
\mathbb{E}[Z^k] &= \mathbb{E}\left[\left(\int_\mu e^{\beta\langle x, Jx\rangle/2}\, dx\right)^k\right] \\
&= \mathbb{E}\left[\int_{\mu^{\otimes k}} \exp\left(\beta\frac{\langle x_1, Jx_1\rangle}{2} + \cdots + \beta\frac{\langle x_k, Jx_k\rangle}{2}\right) dx_1\cdots dx_k\right] \\
&= \int_{\mu^{\otimes k}} \mathbb{E}\left[\exp\left(\beta\frac{\langle \sum_{i=1}^k x_i x_i^T, J\rangle}{2}\right)\right] dx_1\cdots dx_k
\end{aligned}$$

By independence, the expectation factors out, and we can identify each expectation with the MGF of a Gaussian:

$$\mathbb{E}\left[e^{\frac{\beta}{2}J_{ij}\lambda}\right] = e^{\frac{\beta^2}{8n}\lambda^2} \quad \text{for any } \lambda$$

Therefore the integral becomes:

$$\begin{aligned}
\int_{\mu^{\otimes k}} \exp\left(\frac{\beta^2}{4n}\left\|\sum_{i=1}^{k} x_i x_i^T\right\|_F^2\right) &= \int_{\mu^{\otimes k}} \exp\left(\frac{\beta^2}{4n}\operatorname{Tr}\left(\left(\sum_{i=1}^{k} x_i x_i^T\right)^2\right)\right) dx_1 \cdots dx_k \\
&= \int_{\mu^{\otimes k}} \exp\left(\frac{\beta^2}{4n}\sum_{i,j=1}^{k}\langle x_i, x_j\rangle^2\right) dx_1 \cdots dx_k
\end{aligned}$$

Now we can identify $\frac{\langle x_i, x_j\rangle}{n}$ with the overlap matrix Q_{ij} where $Q \in \mathbb{R}^{k\times k}$, and the integral becomes

$$= \int_{\mu^{\otimes k}} \exp\left(\frac{n\beta^2}{4}\sum_{i,j=1}^{k} Q_{ij}^2\right) dx_1 \cdots dx_k = \int_{\mathbb{R}^{k\times k}} \exp\left(\frac{n\beta^2}{4}\sum_{i,j=1}^{k} Q_{ij}^2 + S(Q)\right) \cdot \mathbb{1}_{\left\{Q_{ij} = \frac{\langle x_i, x_j\rangle}{n}\right\}}$$

where

$$S(Q) = \log \int_{\left\{\mu^{\otimes k} \,:\, Q_{ij} = \frac{\langle x_i, x_j\rangle}{n}\right\}} dx_1 \cdots dx_k$$

We now claim that $\lim_{n\to\infty} \frac{S(Q)}{n} = \frac{1}{2}\log\det Q$. We will make the following Gaussian approximation using the "equivalence of ensembles" as before:

$$\operatorname{Unif}\left(\left\{(x_1, \cdots, x_k) \in \mathbb{R}^{nk} \,:\, \frac{\langle x_i, x_j\rangle}{n} = Q_{ij}\right\}\right) \approx \mathcal{N}(0, \Sigma) \quad \text{where} \quad \Sigma = \begin{bmatrix} Q & & \\ & \ddots & \\ & & Q \end{bmatrix} \in \mathbb{R}^{nk\times nk}$$

and $Q_{ij} \in [-1, 1]$. Thus,

$$\begin{aligned}
\log \int_{\left\{\mu^{\otimes k} \,:\, Q_{ij} = \frac{\langle x_i, x_j\rangle}{n}\right\}} 1 \, dx_1 \cdots dx_k &= \log \int_{\left\{\mu^{\otimes k} \,:\, Q_{ij} = \frac{\langle x_i, x_j\rangle}{n}\right\}} e^{k/2} e^{-k/2} \, dx_1 \cdots dx_k \\
&\approx \log \int_{\mathbb{R}^{nk}} e^{nk/2} e^{-\langle y, \Sigma^{-1} y\rangle/2} dy - k \log \mathrm{SA}(\sqrt{n} S^{n-1}) \\
&= n \log \int_{\mathbb{R}^{k}} e^{k/2} e^{-\langle y, Q^{-1} y\rangle/2} \, dy - k \log \mathrm{SA}(\sqrt{n} S^{n-1})
\end{aligned}$$

The surface area term is a normalizing factor from the definition of μ (i.e. normalizing factor integral on $\sqrt{n}S^{n-1}$), and $\frac{nk}{2} \approx \frac{\langle y, \Sigma^{-1} y\rangle}{2}$ by concentration[1]. Continuing on,

$$\begin{aligned}
&= n \log\left((2\pi e)^{k/2} (\det Q)^{1/2}\right) - k \log \mathrm{SA}(\sqrt{n} S^{n-1}) \\
&= \frac{nk}{2}\log(2\pi e) + \frac{n}{2}\log\det Q - k \log \mathrm{SA}(\sqrt{n} S^{n-1}) \\
&= \frac{n}{2}\log\det Q
\end{aligned}$$

[1]This is the same idea as before, by Poincaré's inequality, using that $\Sigma^{-1/2} y$ is a standard Gaussian in nk dimensions.

Therefore,

$$\begin{aligned}
\int_Q \exp\left(\frac{n\beta^2}{4}\sum_{i,j=1}^{k} Q_{ij}^2 + S(Q)\right) &= \int_Q \exp\left(\frac{n\beta^2}{4}\sum_{i,j=1}^{k} Q_{ij}^2 + \frac{nS(Q)}{n}\right) \\
&= \log \max_Q \exp\left(\frac{n\beta^2}{4}\sum_{i,j=1}^{k} Q_{ij}^2 + \frac{nS(Q)}{n}\right) \\
\lim_{n\to\infty} \frac{1}{nk} \log \mathbb{E}[Z^k] &= \frac{1}{k} \max_Q \left\{ \frac{\beta^2}{4}\sum_{i,j=1}^{k} Q_{ij}^2 + \frac{1}{2}\log\det Q \right\}
\end{aligned}$$

1.3 Replica Symmetric Ansatz

Note that $\langle x_i, x_j \rangle = \langle x_j, x_i \rangle$, and the the Gibbs measure concentrates about the top eigenvalue, so a reasonable guess for the Q that maximizes this expression is the **replica symmetric ansatz**:

$$Q = \begin{bmatrix} 1 & q & q & \dots & q \\ q & 1 & q & \dots & q \\ q & q & 1 & \dots & q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q & q & q & \dots & 1 \end{bmatrix} \in \mathbb{R}^{k\times k} \qquad\qquad q \in [0,1]$$

and the expression becomes:

$$\lim_{n\to\infty} \frac{1}{nk} \log \mathbb{E}[Z^k] = \frac{1}{k} \max_{Q,q} \left\{ \frac{k\beta^2}{4} + \frac{k(k-1)\beta^2 q}{4} + \frac{1}{2}\log\det Q \right\}$$

We will continue this computation next lecture by finding q.