Lecture 9: Replica method, II: the spiked GOE matrix, part b)

Lecturer: Song Mei *Scriber: Robbie Netzorg* *Proof reader: Zitong Yang*

1 The Spiked GOE matrix and the free energy approach

In this lecture, we explore using one technique from statistical physics, known as the Replica Method, to help us calculate the asymptotic behavior of the Spiked GOE matrix, given as follows:

Let $\boldsymbol{u} \in \mathbb{S}^{n-1} = \{x \in \mathbb{R}^n \ : \ \|\boldsymbol{x}\|_2 = 1\}$, $\lambda \in R_+$, $\boldsymbol{W} \sim \text{GOE}(n)$, and $Y = \lambda \boldsymbol{u}\boldsymbol{u}^T + \boldsymbol{w} \in \mathbb{R}^{n \times n}$. Then, we are particularly interested in calculating:

$$\phi(\lambda) \equiv \lim_{n\to\infty} \mathbb{E}[\sup_{\boldsymbol{\sigma}\in\mathbb{S}^{n-1}} \langle \boldsymbol{\sigma}, Y\boldsymbol{\sigma}\rangle], \tag{1}$$

$$m(\lambda) \equiv \lim_{n\to\infty} \mathbb{E}[\langle \boldsymbol{v}_{max}(Y), \boldsymbol{u}\rangle^2], \tag{2}$$

where $\boldsymbol{v}_{max}(Y) = \arg\max_{\boldsymbol{\sigma}\in\mathbb{S}^{n-1}} \langle \boldsymbol{\sigma}, Y\boldsymbol{\sigma}\rangle$.

1.1 The Free Energy Approach

[1] Towards the end of calculating the free energy density $\phi(\lambda)$ and the ensemble average of the observable $m(\lambda)$, we introduce a perturbed system given by the Hamiltonian $H_\lambda(\boldsymbol{\sigma})$. With the configuration space $\Omega = \mathbb{S}^{n-1}$ and the reference measure $\nu_0 = \text{Unif}$, we have that

$$\begin{aligned}
H_\lambda(\boldsymbol{\sigma}) \equiv& - n\langle \boldsymbol{\sigma}, \boldsymbol{W}\boldsymbol{\sigma}\rangle - n\lambda\langle \boldsymbol{\sigma}, \boldsymbol{u}\rangle^2, && (3)\\
Z_n(\beta, \lambda) \equiv& \int_{\mathbb{S}^{n-1}} \exp\{-\beta H_\lambda(\boldsymbol{\sigma})\}\nu_0(d\boldsymbol{\sigma}), && (4)\\
\Phi_n(\beta, \lambda) \equiv& \log Z_n(\beta, \lambda), && (5)\\
\phi(\beta, \lambda) \equiv& \lim_{n\to\infty} \mathbb{E}[\log Z_n(\beta, \lambda)]/n, && (6)\\
\phi(\lambda) \equiv& \lim_{\beta\to\infty} \frac{1}{\beta}\phi(\beta, \lambda), && (7)\\
m(\lambda) \equiv& \phi'(\lambda). && (8)
\end{aligned}$$

As is, the above formulas are difficult to work with, especially $\phi(\beta, \lambda)$, which contains the expectation of a logarithm.

2 The Replica Trick

From Sherrington's 1975 work [?], we now introduce the Replica Trick, a helpful technique for simplifying the calculation of $\phi(\beta, \lambda) = \lim_{n\to\infty} \mathbb{E}[\log Z_n(\beta, \lambda)]/n$. For this, we introduce the following Lemma:

Lemma 1. *For a given random variable Z, we have that*

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

[1]For background on this section, please refer to [?]

The usefulness of this lemma lies in its ability to equate the expectation over a logarithm to the expectation over moments, which is much more tractable to analyze.

Applying this lemma to $\phi(\beta, \lambda)$, we have that

$$\phi(\beta, \lambda) = \lim_{k \to 0} \lim_{n \to \infty} \frac{1}{nk} \log \mathbb{E}[Z_n(\beta, \lambda)^k].$$

This gives rise to the following 4 step procedure to calculate $m(\lambda)$:

a) $S(k, \beta, \lambda) = \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}[Z_n^k]$ — The n limit

b) $\phi(\beta, \lambda) = \lim_{k \to 0} \frac{1}{k} S(k, \beta, \lambda)$ — The k limit

c) $\phi(\lambda) = \lim_{\beta \to \infty} \frac{1}{\beta} \phi(\beta, \lambda)$ — The β limit

d) $m(\lambda) = \phi'(\lambda)$ — The λ differentiation

We will now calculate the four steps individually for the Spiked GOE matrix.

3 Calculating the n limit

In order to make the calculation of $S(k, \beta, \lambda)$ tractable, we first introduce the following lemma:

Lemma 2. *For $k \in \mathbb{N}_+$, we have that*

$$\begin{aligned} S(k, \beta, \lambda) &\equiv \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}[Z_n(\beta, \lambda)^k] \\ &= \sup_{\boldsymbol{Q} \in \mathbb{R}^{(k+1)\times(k+1)} \, diag(\boldsymbol{Q}) = 1 \boldsymbol{Q} \succeq 0} U(\boldsymbol{Q}), \end{aligned}$$

where

$$U(\boldsymbol{Q}) = \beta\lambda \sum_{i=1}^{k} q_{0i}^2 + \beta^2 \sum_{i,j=1}^{k} q_{ij}^2 + \frac{1}{2} \log \det(\boldsymbol{Q}),$$

and $\boldsymbol{Q} = (q_{ij})_{0 \le i,j \le k}$.

With this lemma, we have the following derivation:

$$\begin{aligned} \mathbb{E}[Z_n(\beta, \lambda)^k] &= \mathbb{E}\Big[\Big(\int_{\mathbb{S}^{n-1}} \exp\{\beta H_\lambda(\boldsymbol{\sigma})\} \nu_0(d\boldsymbol{\sigma})\Big)^k\Big] \\ &= \mathbb{E}\Big[\int_{(\mathbb{S}^{n-1})^{\otimes k}} \exp\Big\{-\beta \sum_{a=1}^{k} H_\lambda(\boldsymbol{\sigma}^a)\Big\} \prod_{a=1}^{k} \nu_0(d\boldsymbol{\sigma}^a)\Big] \\ &= \int_{(\mathbb{S}^{n-1})^{\otimes k}} \mathbb{E}\Big[\exp\Big\{\beta n \sum_{a=1}^{k} \big(\lambda \langle u, \boldsymbol{\sigma}^a \rangle^2 + \langle \boldsymbol{\sigma}^a, \boldsymbol{W} \sigma^a \rangle\big)\Big\}\Big] \prod_{a=1}^{k} \nu_0(d\boldsymbol{\sigma}^a) \\ &= \int_{(\mathbb{S}^{n-1})^{\otimes k}} \exp\Big\{\beta n \sum_{a=1}^{k} \lambda \langle u, \boldsymbol{\sigma}^a \rangle^2\Big\} \underbrace{\mathbb{E}\Big[\exp\Big\{\beta n \sum_{a=1}^{k} \langle \boldsymbol{\sigma}^a, \boldsymbol{W} \sigma^a \rangle\Big\}\Big]}_{\boldsymbol{E}} \prod_{a=1}^{k} \nu_0(d\boldsymbol{\sigma}^a). \end{aligned}$$

Recalling that $\boldsymbol{W} = (\boldsymbol{G}+\boldsymbol{G}^T)/\sqrt{2n}$, $\boldsymbol{G} \in \mathbb{R}^{n\times n}$, $G_{ij} \sim_{i.i.d.} N(0,1)$, we have from last lecture that

$$\begin{aligned}\boldsymbol{E} = \mathbb{E}\Big[&\exp\Big\{\beta n \sum_{a=1}^{k} \langle \sigma^a, (\boldsymbol{G}+\boldsymbol{G}^T)\sigma^a\rangle/\sqrt{2n}\Big\}\Big] \\ =&\exp\Big\{\beta^2 n \sum_{a,b=1}^{k} \langle \sigma^a, \sigma^b\rangle^2\Big\}.\end{aligned}$$

Substituting $\boldsymbol{E}$, we have that

$$\mathbb{E}[Z_n(\beta,\lambda)^k] = \int_{(\mathbb{S}^{n-1})^{\otimes k}} \exp\Big\{\beta n \sum_{a=1}^{k} \lambda \langle u, \boldsymbol{\sigma}^a\rangle^2 + \beta^2 n \sum_{a,b=1}^{k} \langle \boldsymbol{\sigma}^a, \boldsymbol{\sigma}^b\rangle^2\Big\} \prod_{a=1}^{k} \nu_0(d\boldsymbol{\sigma}^a).$$

Plugging in

$$1 = \int \prod_{a=1}^{k} \delta(\langle \boldsymbol{u}, \boldsymbol{\sigma}^a\rangle - q_{0a}) \prod_{a,b=1}^{k} \delta(\langle \boldsymbol{\sigma}^a, \boldsymbol{\sigma}^b\rangle - q_{ab}) \prod dq_{ab}.$$

We have that

$$\begin{aligned}\mathbb{E}[Z_n(\beta,\lambda)^k] = &\int_{(\mathbb{S}^{n-1})^{\otimes k}} \prod_{0\le a<b\le k}^{k} dq_{ab} \exp\Big\{\beta\lambda n \sum_{a=1}^{k} q_{0a}^2 + \beta^2 n \sum_{a,b=1}^{k} q_{ab}^2\Big\} \\ &\times \underbrace{\int_{(\mathbb{S}^{n-1})+} \prod_{a=1}^{k} \delta(\langle \boldsymbol{u}, \boldsymbol{\sigma}^a\rangle - q_{0a}) \prod_{1\le a<b=\le k}^{k} \delta(\langle \boldsymbol{\sigma}^a, \boldsymbol{\sigma}^b\rangle - q_{ab}) \prod d\nu_0(d\boldsymbol{\sigma}^a)}_{\text{Ent}}.\end{aligned}$$

Using the Laplace Method, we have that

$$\mathbb{E}[Z_n(\beta,\lambda)^k] = \sup_{\substack{\boldsymbol{Q}\succeq 0\\ q_{ii=1}}} \exp\Big\{\beta\lambda n \sum_{a=1}^{k} q_{0a}^2 + \beta^2 n \sum_{a,b=1}^{k} q_{ab}^2\Big\} \times Ent,$$

where $\boldsymbol{Q} = (q_{ij})_{0\le i,j\le k}$.

Simplifying Ent, we have that

$$\begin{aligned}Ent =& \int_{\mathbb{S}^{n-1}} \nu(d\boldsymbol{\sigma}^0) \int_{(\mathbb{S}^{n-1})^k} \prod_{a=1}^{k} \delta(\langle \boldsymbol{\sigma}^0, \boldsymbol{\sigma}^a\rangle - q_{0a}) \prod_{a,b=1}^{k} \delta(\langle \boldsymbol{\sigma}^a, \boldsymbol{\sigma}^b\rangle - q_{ab}) \prod_{a=1}^{k} \nu_0(d\boldsymbol{\sigma}^a) \\ =& \int_{\mathbb{S}^{n-1}(\sqrt{n})} \nu(d\bar{\boldsymbol{\sigma}}^0) \int_{(\mathbb{S}^{n-1}(\sqrt{n}))^k} \prod_{a=1}^{k} \delta(\langle \bar{\boldsymbol{\sigma}}^0, \bar{\boldsymbol{\sigma}}^a\rangle - nq_{0a}) \prod_{a,b=1}^{k} \delta(\langle \bar{\boldsymbol{\sigma}}^a, \bar{\boldsymbol{\sigma}}^b\rangle - nq_{ab}) \prod_{a=1}^{k} \nu_0(d\bar{\boldsymbol{\sigma}}^a) \\ =&\mathbb{P}(\bar{\boldsymbol{Q}}(\bar{\boldsymbol{\sigma}}) \approx \boldsymbol{Q}) = \exp\Big\{n\frac{1}{2}\log\det(\boldsymbol{Q})\Big\}.\end{aligned}$$

Taking everything together, we have that

$$\mathbb{E}[Z_n(\beta,\lambda)^k] = \sup_{\substack{\boldsymbol{Q}\succeq 0\\ \boldsymbol{Q}_{ii}=1}} \exp\Big\{n\Big(\beta\lambda\sum_{a=1}^{k} q_{0a}^2 + \beta^2\sum_{ab=1}^{k} q_{ab}^2 + \frac{1}{2}\log\det(\boldsymbol{Q})\Big\},$$

which implies that

$$\begin{aligned} S(k,\beta,\lambda) &= \sup_{\substack{\boldsymbol{Q}\succeq 0\\ \boldsymbol{Q}_{ii}=1}} U(\boldsymbol{Q}) \\ U(\boldsymbol{Q}) &= \exp\Big\{n\Big(\beta\lambda\sum_{a=1}^{k} q_{0a}^2 + \beta^2\sum_{ab=1}^{k} q_{ab}^2 + \frac{1}{2}\log\det(\boldsymbol{Q})\Big\}. \end{aligned}$$

4 Calculating the k limit

From the previous section, we have that

$$\psi(\beta,\lambda) = \lim_{k\to 0}\frac{1}{k}S(k,\beta,\lambda) = \lim_{k\to 0}\frac{1}{k}\sup_{\substack{\boldsymbol{Q}\succeq 0\\ \boldsymbol{Q}_{ii}=1}} U(\boldsymbol{Q}).$$

There are two problems with the above equation. First, the formula is derived for an integer k. Additionally, k is the dimension of $\boldsymbol{Q}$ matrix, leads to no closed form solution for the argmax.

Our approach is to ignore these problems and guess a form of the solution (ansatz). The hope is that the solution will be correct, which we can verify with simulation studies, and we can return and rigorize the argument.

Defining $\boldsymbol{Q}$ as a $k\times k$ matrix with entries $(q_{ij})_{0\le i,j\le k}$, and $\Pi\in\mathbb{R}^{(k+1)\times(k+1)}$ as

$$\Pi = \begin{bmatrix} 1 & 0 \\ 0 & \bar{\Pi} \end{bmatrix},$$

where $\bar{\Pi}$ is any permutation matrix. We have the following lemma:

Lemma 3. *If $U(\Pi\boldsymbol{Q}\Pi^T) = U(\boldsymbol{Q})$, then there exists some $\boldsymbol{Q}_*\in\mathbb{R}^{(k+1)\times(k+1)}$ that is stationary and $\boldsymbol{Q}_* = \Pi\boldsymbol{Q}_*\Pi^T$ for all permutation matrices $\bar{\Pi}$.*

A reasonable guess is that the supremum is taken at this $\boldsymbol{Q}_*$. Let us posit that $\boldsymbol{Q}$ takes the following form:

$$\boldsymbol{Q} = \begin{bmatrix} 1 & \mu & \dots & \mu \\ \mu & 1 & q & \\ \vdots & q & & \\ \mu & \dots & & 1 \end{bmatrix},$$

where $q_{0a} = \mu$ for all $1\le a\le k$. and $q_{ab} = q$ fpr a;; $1\le a\ne b\le k$. Then, we have the **Replica Symmetric Ansatz**:

$$U(k,\mu,q) = \beta\lambda k\mu^2 + \beta^2(k + k(k-1)q^2) + \frac{1}{2}\log\det(\boldsymbol{Q}).$$

We note that for matrices in the form of $\boldsymbol{Q}$, with a $k\times k$ block $\boldsymbol{A}$, we caluclate its determinant via the Schur Complement:

$$\det \begin{bmatrix} 1 & \mu & \dots & \mu \\ \mu & & & \\ \vdots & & \boldsymbol{A} & \\ \mu & & & \end{bmatrix} = \det(\boldsymbol{A} - \mu^2 1_k 1_k^T),$$

where $\boldsymbol{A} = q11^T + (1-q)\boldsymbol{I}_k$.

Then, we have that

$$\begin{aligned} \log\det(\boldsymbol{Q}) =& \log\det((1-q)\boldsymbol{I}_k + (q-\mu^2)1_k 1_k^T) \\ =& \log\Big(1 + k\frac{(q-\mu^2)}{(1-q)}\Big) + k\log(1-q). \end{aligned}$$

Simplifying $U(k,\mu,q)$, we have that

$$U(k,\mu,q) = \beta\lambda k\mu^2 + \beta^2(k + k(k-1)q^2) + \frac{1}{2}\Big[k\log(1-q) + \log\Big(1 + k\frac{(q-\mu^2)}{(1-q)}\Big)\Big].$$

With our guess, we have the following:

$$\begin{aligned} \phi(\beta,\lambda) =& \lim_{k\to 0}\sup_{\mu,q}\frac{1}{k}U(k,\mu,q) \\ =& \mathrm{ext}_{\mu,q}\lim_{k\to 0}\frac{1}{k}U(k,\mu,q), \end{aligned}$$

where $\mathrm{ext}_x f(x) = f(x_*)$ where $x_* \in \{x \ : \ \nabla f(x) = 0\}$. Then we have that

$$\begin{aligned} u(\mu,q;\beta,\lambda) =& \lim_{k\to 0}\frac{1}{k}U(k,\mu,q) \\ =& \beta\lambda\mu^2 + \beta^2(1-q^2) + \frac{1}{2}\log(1-q) + \frac{1}{2}\frac{q-\mu^2}{1-q}, \end{aligned}$$

and

$$\phi(\beta,\lambda) = \mathrm{ext}_{\mu,q} u(\mu,q;\beta,\lambda).$$

Per the definition of the ext operator, we solve for μ, q s.t. the partial derivatives are equal to 0:

$$\begin{aligned} \partial_\mu u =& 2\beta\lambda\mu - \frac{\mu}{1-q} = \mu\Big(2\beta\lambda - \frac{1}{1-q}\Big) \\ \partial_q u =& -2\beta^2 q - \frac{1}{2(1-q)} + \frac{1}{2}\frac{q}{1-q} + \frac{1}{2}\frac{q-\mu^2}{(1-q)^2} \\ =& -2\beta^2 q - \frac{\mu^2}{2(1-q)^2} + \frac{1}{2(1-q)^2}. \end{aligned}$$

Solving these equations, we have the following extrema:

1. $\mu_1 = 0$, $q_1 = 1 - 1/2\beta$, which implies that

$$\phi_1(\beta,\lambda) = 2\beta - \frac{3}{4} - \frac{1}{2}\log(2\beta).$$

2. $2\beta\lambda = 1/(1-q)$, which implies that

$$\mu_2 = \Big(\Big(1 - \frac{1}{\lambda^2}\Big)\Big(1 - \frac{1}{2\beta\lambda}\Big)\Big)^{1/2},$$
$$q_2 = 1 - \frac{1}{2\beta\lambda}.$$

Additionally, we have that

$$\phi_2(\beta,\lambda) = \beta\Big(\lambda + \frac{1}{\lambda}\Big) - \Big(\frac{1}{4\lambda^2} + \frac{1}{2}\Big) - \frac{1}{2}\log(2\lambda\beta).$$

5 Computing the $\beta \to 0$ Limit

We are close to having computed the free energy density, $\phi(\lambda)$. All that remains is taking the limit with respect to β. With that in mind, computing the limits, we have that

$$\phi_1(\lambda) = \lim_{\beta\to\infty} \frac{1}{\beta}\phi_1(\beta,\lambda) = 2,$$
$$\phi_2(\lambda) = \lambda + \frac{1}{\lambda}.$$

Noting that $\|\boldsymbol{W}\|_{op} \approx 2$ and $\|\lambda \boldsymbol{u}\boldsymbol{u}^T\|_{op} \approx \lambda$, we note that

$$\phi(\lambda) = \lim_{n\to\infty} \mathbb{E}[\lambda_{max}(\lambda \boldsymbol{u}\boldsymbol{u}^T + \boldsymbol{W})]$$

has two properties:

1. $\phi(\lambda)$ is non-decreasing
2. $\lim_{n\to\infty} \phi(\lambda) = \infty$

Plotting ϕ_1, ϕ_2 in Figure 1, we see that:

1. $\lambda \le 1$, ϕ_2 is decreasing, which implies that the solution is $\phi_1 = 2$
2. $\lambda < 1$, ϕ_1 stays constant, which implies that the solution is ϕ_2

This defines the BBP transition:

$$\phi(\lambda) = \begin{cases} 2 & \lambda \le 1 \\ \lambda + \frac{1}{\lambda} & \lambda > 1 \end{cases}.$$

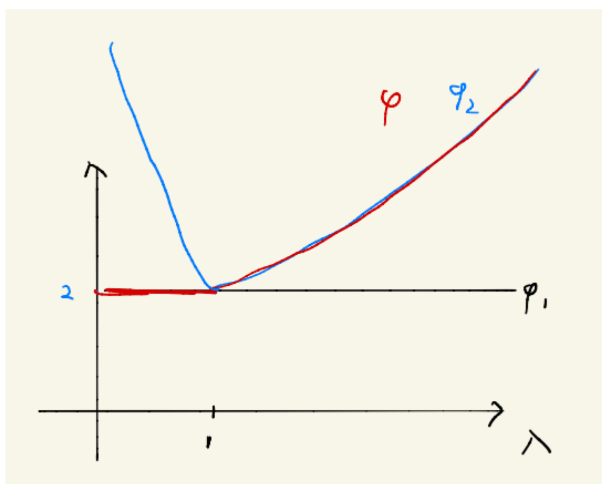

Figure 1: BBP Transition Illustrated