CPSC 664, Spring 2023 Scribe: Peixin You
March 30, 2023 Lecturer: Tim Kunisky

Lecture 18: Replica analysis of SK model

1 Problem Statement

Suppose we have a normalized GOE matrix $W \sim GOE(N, 1/N)$ (so that the empirical spectral distribution converges to the semicircle distribution from -2 to 2) and we have the objective function

$$H(x) = x^T W x \text{ over } x \in \{\pm 1\}^N.$$

In this lecture, we will try to use replica analysis to solve this question.

The usual physics setup is as follows: we consider the Gibbs measure

$$\mu_\beta(x) = \frac{1}{Z(\beta) \exp(\beta H(x))}$$

which is the probability distribution over the hypercube, when β goes bigger it weights more and more the objective function. We define the related

$$Z_W(\beta) = \sum_x \exp(\beta H(x)), \qquad F_W(\beta) = \log(Z(\beta)).$$

From last time, we know F_W will scale with N so we want to catch that number by defining:

$$f(\beta) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_W F_W(\beta)$$

We can show

$$2P_* := \lim_{N\to\infty} \frac{1}{N} \mathbb{E} \max_x H(x) = \lim_{\beta\to\infty} \frac{1}{\beta} f(\beta),$$

and we will try to make a prediction for $f(\beta)$ to compute the right-hand side.

2 Idea of Replica trick

The idea of the replica trick is that

$$\mathbb{E} \log Z = \lim_{n\to 0} \frac{\log \mathbb{E} Z^n}{n}$$

So we could move the expectation inside the log. We want to apply the replica method on $\mathbb{E}_W F_W(\beta)$. We have

$$\begin{aligned}
\mathbb{E}F(\beta) &= \mathbb{E} \log Z(\beta) \\
&= \lim_{n\to 0} \frac{\log \mathbb{E} Z(\beta)^n}{n}
\end{aligned}$$

Also notice that

$$\mathbb{E}_W Z(\beta)^n = \mathbb{E}_W \sum_{x_1,\ldots,x_n} \mathsf{exp}\left(\beta \sum_{i=1}^n H(x_i)\right)$$
$$= \sum_{x_1,\ldots,x_n} \mathbb{E}_W \,\mathsf{exp}\left(\beta \sum_{i=1}^n H(x_i)\right)$$

Now we want to apply the following identity.

Theorem 2.1. $\mathbb{E}_{W\sim GOE(N,\sigma^2)} \,\mathsf{exp}\,(\langle W, A\rangle) = \mathsf{exp}\,(\sigma^2 \|A\|_F^2)$.

We rewrite $\mathsf{exp}(\beta \sum_{i=1}^n H(x_i))$ to $\langle W, \beta \sum x_i x_i^T\rangle$. For convenience, let $X = [x_1, \ldots, x_n]$, then $\sum x_i x_i^T = XX^T$.

So now we have

$$\mathbb{E}_W Z(\beta)^n = \sum_{x_1,\ldots,x_n} \mathsf{exp}\left(\frac{1}{N}\|\beta X X^T\|_F^2\right)$$
$$= \sum_{x_1,\ldots,x_n} \mathsf{exp}\left(N\beta^2 \left\|\frac{1}{N} X^T X\right\|_F^2\right).$$

Let $Q(x_1,\ldots,x_n) := \frac{1}{N}X^TX$, so $Q(x_1,\ldots,x_n)_{i,j} = \frac{1}{N}\langle x_i, x_j\rangle \in [-1,1]$. In particular, $Q \succeq 0$ and $Q_{i,i} = 1$.

We then have

$$\mathbb{E}_W Z(\beta)^n = \sum_{Q \text{ achievable}} \mathsf{exp}\,(N\beta^2 \|Q\|_F^2) \cdot \#\{(x_1,\ldots,x_n) \in (\{\pm 1\}^N)^n : Q(x_1,\ldots,x_n) = Q\}.$$

We can see

$$\#\{(x_1,\ldots,x_n) \in (\{\pm 1\}^N)^n : Q(x_1,\ldots,x_n) = Q\} = 2^{nN}\mathbb{P}_{x_1,\ldots,x_n\sim \mathrm{Unif}(\{\pm -1\}^N)}[Q(x_1,\ldots,x_n) = Q]$$

Then, by a Laplace method approximation,

$$\mathbb{E}_W Z(\beta)^n \approx \mathsf{exp}\left(N \sup_{Q\succeq 0, Q_{ii=1}} \left\{n \,\mathsf{log}\, 2 + \beta^2 ||Q||_F^2 + \frac{1}{N}\,\mathsf{log}\, \mathbb{P}[Q(x_1,\ldots,x_n) \approx Q]\right\}\right)$$

We give a heuristic for computing the latter probability. Let y_i be the ith row of X. so we could rewrite Q to $\frac{1}{N}\sum y_i y_i^T$. Then, we have

$$\mathbb{P}[Q(x_1,\ldots,x_n) \approx Q] = \mathbb{P}\left[\frac{1}{N}\sum_1^N y_i y_i^T \approx Q\right]$$

This should be viewed as a *large deviations* probability. Following scalar large deviations theory, we consider a Chernoff-type bound parametrized by a matrix Λ,

$$\begin{aligned}
&\leq \mathbb{P}[\mathsf{exp}\langle\Lambda, \sum y_i y_i^T\rangle \geq \mathsf{exp}\langle\Lambda, NQ\rangle] \\
&\leq \frac{\mathbb{E}\,\mathsf{exp}(\langle\Lambda, \sum y_i y_i^T)}{\mathsf{exp}\,N\langle\Lambda, Q\rangle} \\
&= \left(\frac{\mathbb{E}_y\,\mathsf{exp}\,y^T\Lambda y}{\mathsf{exp}\langle\Lambda, Q\rangle}\right)^N \\
&:= B(\Lambda)^N.
\end{aligned}$$

As in the scalar case, one can show (in a version of *Cramér's theorem*) that

$$\mathbb{P}[Q(x_1, \ldots, x_n) \approx Q] \approx (\inf_\Lambda B(\Lambda))^N.$$

Now we have

$$\mathbb{E}Z(\beta)^n \approx \mathsf{exp}\left(N \sup_Q \inf_\Lambda \left\{n \log 2 + \beta^2 ||Q||_F^2 + \log \mathbb{E}_y \,\mathsf{exp}(y^T\Lambda y) - \langle\Lambda, Q\rangle\right\}\right).$$

If we don't have the restriction of y in the hypercube but in the full sphere, we could compute the $\mathbb{E}_y$ term explicitly. But in our case, this term creates many complications. As before, our approach will be to make a guess for Q, which will lead us to also make a guess for Λ and simplify this expression to something more tractable.

3 Replica Symmetric ansatz [SK]

We will assume the optimizer $Q^\star$ looks like

$$Q^* = \begin{pmatrix} 1 & a & \cdots & a \\ a & 1 & \cdots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \cdots & 1 \end{pmatrix}.$$

This means the diagonal is 1, and all other place is constant a. We may think of this as the Gram matrix describing the typical relative configuration of n independent draws from the Gibbs measure μ_β. For example, when $\beta = 0$, these draws will be uniform in the hypercube and we will find that this "overlap matrix" should look like the identity.

What we are asking is how Q^* is going to change when we increase β, which means we care more about our objective function. Basically, the above assumption says inside the hypercube, there are some regions where all the good solutions are and which look like "geometrically simple" lower-dimensional hypercubes or spheres, and the Gibbs measure is uniform over those regions. We will see that this guess will ultimately be wrong, but we will follow the historical development and give these details first.

In order to approximate our $\mathbb{E}Z(\beta)^n$, we also want to know what Λ could be under the assumption of our optimal Q. Function $h(\Lambda) = \log \mathbb{E}\,\mathsf{exp}\,y^T\Lambda y$ is convex, so if we replace function Λ with $\bar{\Lambda}_{i,i}$ be the average of the diagonal entry which is $\frac{1}{n}\sum \Lambda_{i,i}$ and $\bar{\Lambda}_{i,j} = \frac{1}{n(n-1)}\sum_{k\neq l}\Lambda_{k,l}$, then we can show $h(\bar{\Lambda}) \leq h(\Lambda)$ and $\langle \Lambda, Q\rangle = \langle \bar{\Lambda}, Q\rangle$. So we can assume:

$$\Lambda^\star = \begin{pmatrix} c & b & \cdots & b \\ b & c & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & c. \end{pmatrix}$$

If we look at the value of the diagonal of $\mathbb{E}\exp y^T\Lambda y$ and $\langle \lambda, Q\rangle$, both of them are equal $\sum \Lambda_{i,i}$ and they have different signs in the objective function, which means the value of the diagonal won't affect the result. So, we may assume $c = b$ so that $\Lambda^\star = b11^\top$.

Then, we have

$$\begin{aligned} h(\Lambda^\star) &= \log \mathbb{E}_y \exp(b(\sum y_i)^2) \\ &= \log \mathbb{E}_{y\sim \text{ Unif}(\{\pm 1\}^n, g\sim N(0,1))} \exp(g\sqrt{2b}\sum y_i) \\ &= \log \mathbb{E}_g(\mathbb{E}_{y\sim \text{Unif}(\{\pm 1\})} \exp(g\sqrt{2b}y))^n \\ &= \log \mathbb{E}_g(\cosh g\sqrt{2b})^n \end{aligned}$$

So, we finally find

$$\mathbb{E}Z(\beta)^n \approx \exp\left(N \sup_a \inf_b \left\{n\log 2 + \beta^2(n + n(n-1)a^2) + \mathbb{E}_g \cosh(g\sqrt{2b})^n - nb - abn(n-1)\right\}\right).$$

We will continue the calculation from here in the next lecture.