Lecture 17: Introduction to Sherrington-Kirkpatrick (SK) model

1 Problem Definition

This is a very important model in statistical physics and we will motivate this problem with a pure combinatorial problem, the max cut problem.

We've seen the following in Homework 1

$$W \sim GOE(N, \frac{1}{N})$$

$||W|| = O(1)$ w.h.p., then we define the following optimization problem

$$M(W) = \max_{x \in \{\pm 1\}^N} x^T W x$$

$||x|| = \sqrt{N}$ w.h.p., this is like the eigenvalue problem but over the hypercube, which changes the problem drastically and this constraint makes it harder but more interesting.

2 Motivation

This problem is motivated by a combinatorial problem, the MaxCut problem, specified by a graph $G = (V, E)$. We devide the set V into two sets and maximize the number of edges between these two sets. This is a well-known NP-complete problem.

Claim 1. *The normalized MaxCut is a special case of our optimization problem M.*

Proof. Given a G, we have two matrices: $A = A_G$ the adjacency matrix and a diagonal matrix $D = D_G$ where $D_{xx} = deg_G(x)$. Then we can build the Laplacian matrix $L = L_G = D - A$. If we look at the quadratic form:

$$\begin{aligned} v^T L v &= \sum_x deg_G(x) v_x^2 - 2 \sum_{(x,y) \in E} v_x v_y \\ &= \sum_{(x,y) \in E} (v_x^2 + v_y^2) - 2 \sum_{(x,y) \in E} v_x v_y \\ &= \sum_{(x,y) \in E} (v_x - v_y)^2 \end{aligned}$$

which can be regarded as the "gradient" of v: $||\nabla_v||^2$. If we plug in x in the hypercube so that $x \in \{\pm 1\}^V$ corresponds to a division of the vertices set $U \subset V$, then $x^T L x = 4 \times \{\#\text{ of edges cut by } U\}$. So $MaxCut(G) = M(\frac{1}{4}L)$.

□

Claim 2. *Plugging in these "smooth" Gaussian matrices W corresponds to meaningful Max-Cut problem.*

Let's look at MaxCut for random graphs.
For $d \geq 3$, $G \sim Unif(\{\text{d-regular graph on N vertices}\})$.

Remark 2.1. *One can utilize configuration models to efficiently sample from this distribution.*

Since d is independent from N, G is a sparse graph: $|E| = \frac{dN}{2} = O(N)$. One could ask what is the typical MaxCut value for this graph.
Since it's $d-regular$, $D = dI$ so $x^T L x = dN - x^T A x$ and:

$$MaxCut(G) = \left[\frac{1}{4}dN - \frac{1}{4}\min_x x^T A x\right]$$
$$Normalized\_MaxCut(G) = \left[\frac{1}{2} - \frac{1}{2dN}\min_x x^T A x\right]$$

For simplicity we will use $MaxCut$ to denote $Normalized\_MaxCut$.

3 Bounds for MaxCut

3.1 Upper Bound

We can upper bound it with

$$\begin{aligned} MaxCut(G) &\leq \left[\frac{1}{2} - \frac{1}{2dN}\lambda_{min}(A)||x||^2\right] \\ &= \left[\frac{1}{2} - \frac{1}{2dN}\lambda_{min}(A)N\right] \\ &= \left[\frac{1}{2} + \frac{||\lambda_{min}(A)||}{2d}\right] \end{aligned}$$

Let's denote the eigenvalues of A as: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, then $\lambda_1 = d$ with eigenvector being all 1's, which is evidence that the results about the empirical spectral distribution from the free probability tool do not describe the outliers of the eigenvalues. ($d > 2\sqrt{d-1}$)

Theorem 3.1. *W.h.p., $-2\sqrt{d-1} - o_{N\to\infty}(1) \leq \lambda_N \leq \lambda_1 \leq 2\sqrt{d-1} + o_{N\to\infty}(1)$*

And by the above theorem, one can show that w.h.p.:

$$MaxCut(G) \leq \frac{1}{2} + \frac{\sqrt{d-1}}{d} + o(1)$$

3.2 Lower Bound

We can also lower bound this by analyzing some "local algorithms" that given G, output a cut. W.h.p.:

$$MaxCut(G) \geq \frac{1}{2} + \frac{c}{\sqrt{d}}$$

.

3.3 Problems to Study

- **Q1**: For small constants d (i.e. $d = 3$), find $\lim_N \mathbb{E} MaxCut(G)$
- **Q2**: For large d, find the threshold.

We focus on the second problem. Specifically, we are interested in showing the existence of such constant c, so that $\lim_N \mathbb{E} MaxCut(G) = \frac{1}{2} + \frac{c}{\sqrt{d}} + o(\frac{1}{\sqrt{d}})$ as $d \to \infty$. And if it exists, we want to find out what it is.

Remark 3.2. *Same for sparse ErdosRenyi graphs (edges are present with probability $\frac{d}{N}$). The same argument does not apply but the same conclusion can be drawn.*

Remark 3.3. *Same for minimum bisection (i.e. $||U|| = \frac{N}{2}$) and maximum bisection problem. In simulations, it is close to MaxCut.*

Conjecture 3.4. $\lim_{N\to\infty} MaxCut(G) - \frac{1}{2} = \frac{1}{2} - \lim_{N\to\infty} MinBis(G)$ *even for fixed d.*

4 Find Constant c

Recall that we have:

$$\frac{1}{2} - \frac{1}{2dN} \min_x x^T A x = MaxCut(G) = \frac{1}{2} + \frac{c}{\sqrt{d}} + o(\frac{1}{\sqrt{d}})$$

It's equivalent to that w.h.p.:

$$\frac{1}{2\sqrt{d}N} \min_x x^T A x = -c + o(1)$$
$$\Leftrightarrow \frac{1}{2N} \min_x x^T \frac{A}{\sqrt{d}} x = -c + o(1)$$

Recall that "free CLT" says that e.s.d. of $\frac{A}{\sqrt{d}}$ is the semi-circle, which is the e.s.d. of $GOE(N, \frac{1}{N})$, so heuristically

$$
\begin{aligned}
\frac{1}{2N}\min_x x^T\frac{A}{\sqrt{d}}x \approx \frac{1}{2N}\min_x x^T W x \\
&= -\frac{1}{2N}\max_x x^T(-W)x \\
&= -\frac{1}{2N}\max_x x^T W x
\end{aligned}
$$

Heuristic guess of the constant:

$$c = \lim_{N\to\infty}\frac{1}{2N}\mathbb{E}_{W\sim GOE(N,\frac{1}{N})}M(W)$$

, which is the ground state of this SK model.

Theorem 4.1. *The Parisi number P^* (i.e. the above c value) exists and has some complicated formula. The number is 0.7632.*

Theorem 4.2. $\lim_{N\to\infty}\mathbb{E}MaxCut(G) = \frac{1}{2} + \frac{P^*}{\sqrt{d}} + o(\frac{1}{\sqrt{d}})$

5 Replica Method for P^*

In physics, the SK Hamiltonian $H(x) = x^T W x$. Then we can define:

- $\mu_\beta(x) = \frac{1}{Z(\beta)}e^{\beta H(x)}$, where $Z(\beta) = \sum_x e^{\beta H(x)}$
- Free energy: $F(\beta) = \log(Z(\beta))$; Free energy density: $f(\beta) = \lim_{N\to\infty}\frac{1}{N}F(\beta)$
- Ground state: $\lim_{\beta\to\infty}\frac{1}{\beta}f(\beta)$

Note that $\mathbb{E}F(\beta) = \mathbb{E}\log Z(\beta) = \lim_{n\to 0}\frac{\mathbb{E}Z(\beta)^n}{n}$
while $\mathbb{E}Z(\beta)^n = \sum_{x_1,\cdots,x_n}\mathbb{E}_W e^{\beta(\langle W, x_1x_1^T+\cdots+x_nx_n^T\rangle)}$.

The exponent could be written as the Frobenius norm of XX^T and then what we did previously is to transform this to the Frobenius norm of X^TX, which is a small matrix. And in the end one could get some results by making some naive assumption that the matrix has some certain number on the diagonal and some other number on the off-diagonal entries. But one would find out that this result would imply some "negative entropy", which could not happen. We will see in the next lectures that some replica-symmetry-breaking could be used to repair this technique to calculate the P^* value.