CSE 525: Randomized Algorithms **Spring 2026**

Lecture 18: Hypergraph Sparsification

Lecturer: Shayan Oveis Gharan *06/02/26*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

18.1 Max Eigenvalue of Random Matrices

We will now use generic chaining to show that the largest eigenvalue of a random symmetric $n \times n$ matrix with Rademacher entries is $O(\sqrt{n})$. This is certainly not the simplest way of proving such a result, but it will give a sense of how these techniques can be applied.

Our "sub-Gaussian" random process is to pick Rademacher random variables $r_{i,j}$, for each $1 \leq i < j \leq n$, define the matrix $M_{i,j} = M_{j,i} = r_{i,j}$ and let

$$F(x) = x^T M x$$

for every $x \in T$. (we assume the diagonal is 0 for simplicity. It thus follows that

$$x^T M x = 2 \sum_{1 \leq i < j \leq n} x_i x_j r_{i,j} = \langle r, z \rangle$$

So, we let T be the set of vectors

$$T = \{z \in \mathbb{R}^{\binom{n}{2} - n} : \forall i, j : z_{\{i,j\}} = 2x_i, x_j, \|x\| = 1\}$$

Let g be a Radamacher random variable By Hoeffding's inequality for any vector v we can write

$$\mathbb{P}\left[\langle g, v \rangle \geq \ell\right] \leq \exp\left(-\frac{\ell^2}{\sum_i 4v_i^2}\right) = \exp(-\frac{\ell^2}{\|v\|^2}),$$

so, this distribution is 1/4-subgaussian.

It follows that for $x, y \in \mathbb{R}^n$ with $\|x\| = \|y\| = 1$

$$\begin{aligned}
\frac{1}{4}\|z^x - z^y\|^2 \leq \|xx^T - yy^T\|_F^2 &= \|(x-y)x^T - y(x-y)^T\|_F^2 \\
&\leq \|(x-y)x^T\|_F^2 + \|y(x-y)^T\|_F^2 \\
&= \|x-y\|^2\|x\|^2 + \|x-y\|^2\|y\|^2 \\
&\leq 2\|x-y\|^2.
\end{aligned}$$

Having this we can write

$$d(z^x, z^y)^2 = \mathbb{E}\left[\langle g, z^x - z^y \rangle^2\right] = \|z^x - z^y\|^2 \leq 8\|x-y\|^2.$$

Now we need to apply generic chaining to F. We can conclude that an ϵ-net over the unit Euclidean sphere is also a $\sqrt{8} \cdot \epsilon$-net for the metric space (T, d). For the unit Euclidean sphere there is an ϵ-net of size at most

$(3/\epsilon)^n$. To apply generic chaining, let T_k be an arbitrary subset of T of cardinality 2^{2^k} if $2^k < n$, and an ϵ-net with $\epsilon = 3 \cdot 2^{-2^k/n}$ otherwise. Applying the generic chaining inequality,

$$\mathbb{E}\left[\sup_{z \in T} F(z)\right] \leq O(1) \cdot \sum_k 2^{k/2} \cdot \sqrt{8} \cdot \min\left\{2, 3 \cdot 2^{-2^k/n}\right\} = O(\sqrt{n})$$

Consider a weighted hypergraph $H = (V, E, w)$ where $w_e : e \in E$ are nonnegative edge weights. We associated to H the quadratic expression

$$Q_H(x) = \sum_{e \in E} w_e \max_{u,v \in e} (x_u - x_v)^2.$$

The main observation is that If H were a graph i.e., for every edge e we had $|e| = 2$ for, this would correspond to the quadratic form of the graph Laplacian.

As our main application of chaining we will explain algorithms to sparsify hypergraphs: That is we want to construct another hypergraph $\tilde{H} = (V, \tilde{E}, \tilde{w})$ such that $\tilde{E} \subseteq E$ and such that

$$|Q_H(x) - Q_{\tilde{H}}(x)| \leq \epsilon \cdot QH(x), \qquad \forall x \in \mathbb{R}^n \tag{18.1}$$

where as usual $\epsilon > 0$ is the accuracy parameter of our sparsifier. Furthermore, similar to the graph case we would like to make $|\tilde{E}|$ as small as possible, ideally near-linear in $|V|$ (while —E— could be as large as $2^{|V|}$ in this case).

The following theorem is proved in a paper by James Lee

Theorem 18.1. *For any n-vertex weighted hypergraph $H = (V, E, w)$ and $\epsilon > 0$, there is a spectral ϵ-sparsifier $\tilde{H} = (V, \tilde{E}, \tilde{w})$ for H such that*

$$|\tilde{E}| \leq O(\frac{\log D}{\epsilon^2} n \log n),$$

where $D := \max_{e \in D} |e|$.

18.2 Independent random sampling

For any edge e let Q_e defined as

$$Q_e(x) = w_e \max_{u,v \in E} (x_u - x_v)^2.$$

Therefore, $Q_H(x) = \sum_e Q_e(x)$.

Suppose we have a probability distribution $p_e : e \in E$, i.e., $p_e \geq 0$ and $\sum_e p_e = 1$. Similar to the graphic case, we let X be an unbiased estimator: Namely we let $X = \frac{w_e}{p_e} Q_e$ with probability p_e. Then, it follows that

$$\mathbb{E}[X] = \sum_e p_e \cdot \frac{w_e}{p_e} Q_e = Q_H$$

As usual the difficulty would be in choosing the probabilities p_e.

Then, we form $\tilde{H}$ by sampling X, m many times and taking the empricial mean of the samples. In particular, we can write,

$$Q_{\tilde{H}}(x) = \frac{1}{m}\sum_{k=1}^{m} \frac{w_{e_k}}{p_{e_k}} Q_{e_k}(x),$$

Observe that

$$\mathbb{E}\left[Q_{\tilde{H}}(x)\right] = Q_H(x)$$

for all $x \in \mathbb{R}^n$. So, as before the main question is how to choose the probabilities p_e?

18.3 Auxilury Graph

Define the edge set

$$F := \bigcup_{e \in E} \binom{e}{2},$$

and let $G = (V, F, c)$ be a weighted graph, where we will choose the edge conductances $c \in \mathbb{R}_F^+$ later. Let

$$L_G := \sum_{\{u,v\} \in F} c_{uv}(b_{u,v})(b_{u,v})^T,$$

Let $R(u, v)$ denote the effective resistance between u, v in G. For a hyperedge $e \in H$, we let

$$R_{\max}(e) := \max\left\{R_{u,v} : \{u, v\} \in e\right\},$$

Having this we define

$$p_e := \frac{w_e R_{\max}(e)}{Z}, \quad \text{for } e \in E.$$

where $Z := \sum_{e \in E} w_e R_{\max}(e)$ is the normalizing constant. Note that in the special case that H is a graph $R_{\max}(e) = R(e)$. Now, the question is how to choose the conductances of the edges of G?

The following is the main lemma:

Lemma 18.2. *Suppose it holds that*

$$\|x\|^2 \leq Q_H(L_G^{-1/2}x), \forall x \in \mathbb{R}^n, \tag{18.2}$$

then for any $1 > \epsilon > 0$ *and* $m \leq O(\frac{\log D}{\epsilon^2} Z \log n)$*, with a constant probability* $\tilde{H}$ *is a sparsifier of* H*.*

The proof of this lemma uses the chaining machinery. But, let us first discuss how to satisfy assumptions of this lemma?

Roughly speaking this auxiluary graph G, puts Q_H into isotropic position. Of course, this step is very straightforward for matrices but as you will see this is fairly more complicated for these non-linear operators.

18.4 Choosing Conductances

We are therefore left to find edge conductances in the graph $G = (V, F, c)$ so that (18.2) holds and Z is small. To this end, let us choose nonnegative numbers

$$\{c_{uv}^e \geq 0 : \{u, v\} \in \tbinom{e}{2}, e \in E\}$$

such that

$$\sum_{\{i,j\}\in\binom{e}{2}} c_{uv}^e = w_e, \quad \forall e \in E. \tag{3.11}$$

For $\{u, v\} \in F$, we then define our edge conductance

$$c_{uv} := \sum_{e\in E:\{u,v\}\in\binom{e}{2}} c_{uv}^e. \tag{18.3}$$

In this case,

$$\begin{aligned}\|L_G^{1/2}y\|^2 = \langle y, L_G y\rangle = \sum_{\{i,j\}\in F} c_{uv}(y_u - y_v)^2 &= \sum_{e\in E}\sum_{\{u,v\}\in\binom{e}{2}} c_{uv}^e (y_u - y_v)^2 \\ &\le \sum_{e\in E}\sum_{\{u,v\}\in\binom{e}{2}} c_{uv}^e \max_{\{u,v\}\in\binom{e}{2}} (y_u - y_v)^2 \\ &= \sum_{e\in E} w_e \max_{\{i,j\}\in\binom{e}{2}} (y_u - y_v)^2 = Q_H(y).\end{aligned}$$

Taking $y = L_G^{+1/2}x$ gives

$$\|x\|^2 \le Q_H(L_G^{+1/2}x),$$

verifying (18.2).

Lemma 18.3 (Foster's Network Theorem)**.** *It holds that*

$$\sum_{\{u,v\}\in F} c_{uv} R(u, v) = n - 1.$$

Proof. The observation is that

$$\sum_{u,v} c_{uv} R(u, v) = \sum_{u,v} c_{uv} \operatorname{Tr}(b_{u,v}^T L_G^{-1} b_{u,v}) = \sum_{u,v} c_{uv} \operatorname{Tr}(b_{u,v} b_{u,v}^T L_G^{-1}) = \operatorname{Tr}(L_G^{-1} L_G) = n - 1.$$

□

Now, define

$$K := \max_{e\in E} \max_{u,v\in\binom{e}{2}} \frac{R_{\max}(e)}{R(u,v)} \mathbb{I}\left[c_{u,v}^e > 0\right].$$

Then, $Z = \sum_{e\in E} w_e R_{\max}(e) \le K(n - 1)$.

Lemma 18.4. *We can choose conductances $c_{u,v}^e$ such that (18.3) is satisfied and $K = 1$.*

Proof. The conductances can be computed by solving the following convex program:

$$\begin{aligned} \max \quad & \log\det(L_G(c_{u,v}) + J) \\ \text{s.t.,} \quad & \sum_{u,v\in\binom{e}{2}} c_{u,v}^e = w_e \\ & c_{u,v}^e \ge 0. \end{aligned} \tag{18.4}$$

Note that this program is convex as log of the determinant is a concave function. Equivalently, the objective can be written as log of the generating polynomial of all spanning trees of G: $\log \sum_T \prod_{\{u,v\}\in T} c_{u,v}$. The convexity follows by the fact that any real stable polynomials is log-concave.

We don't go into the details here. The proof uses writes the "KKT" condition for the optimality of the convex program and deduces the bound on K from that. We remark that □

18.5 Notes on Proof of Lemma 18.2

The proof of Lemma 18.2 is technical but at high-level it uses the generic chaining machinery. The set $T = \{x : Q_H(L_G^{-1/2}x) \leq 1\}$ and the proof uses (18.2) which says that T is a subset of the unit ball.

We just explain the first few steps: We let $\hat{H}$ be an independent copy of $\tilde{H}$,

$$\begin{aligned}\mathbb{E}_{\tilde{H}} \max_{Q_H(y)\leq 1} |Q_H(y) - Q_{\tilde{H}}(y)| &= \mathbb{E}_{\tilde{H}} \max_{Q_H(y)\leq 1} |\mathbb{E}_{\hat{H}}[Q_{\hat{H}}(y)] - Q_{\tilde{H}}(y)| \\ &= \mathbb{E}_{\tilde{H}} \max_{Q_H(y)\leq 1} |\mathbb{E}_{\hat{H}}[Q_{\hat{H}}(y) - Q_{\tilde{H}}(y)]| \\ &\leq \mathbb{E}_{\hat{H},\tilde{H}} \max_{Q_H(y)\leq 1} |Q_{\hat{H}}(y) - Q_{\tilde{H}}(y)|,\end{aligned}$$

where we have used that $|\mathbb{E}X| \leq \mathbb{E}|X|$ and $\max(\mathbb{E}X_1, \ldots, \mathbb{E}X_k) \leq \mathbb{E}\max(X_1, \ldots, X_k)$.

The point of this first few steps is to make the process "centered". The second step is to avoid binary 0/1 random variables. Such random variables are very annoying to run a chaining argument for.

The idea is that the distribution of $Q_{\hat{e}_k}(y) - Q_{\tilde{e}_k}(y)$ is symmetric around the origin, i.e., centered. So, $Q_{\hat{e}_k}(y) - Q_{\tilde{e}_k}(y)$ and $-(Q_{\hat{e}_k}(y) - Q_{\tilde{e}_k}(y))$ have the same distribution.

$$\begin{aligned}\mathbb{E}_{\hat{H},\tilde{H}} \max_{Q_H(y)\leq 1} |Q_{\hat{H}}(y) - Q_{\tilde{H}}(y)| &\leq \mathbb{E}_{s\in\{-1,+1\}^m}\mathbb{E}_{\hat{H},\tilde{H}} \max_{Q_H(y)\leq 1} \left| \frac{1}{m}\sum_{k=1}^m s_i \left(\frac{w_{\hat{e}_i}}{p_{e_i}} Q_{\hat{e}_i}(y) - \frac{w_{\tilde{e}_i}}{p_{\tilde{e}_i}} Q_{\tilde{e}_i}(y) \right) \right| \\ &\leq 2\mathbb{E}_{s\in\{-1,+1\}^m}\mathbb{E}_{\hat{H}} \max_{Q_H(y)\leq 1} \left| \frac{1}{m}\sum_{k=1}^m s_i \frac{w_{\hat{e}_i}}{p_{e_i}} Q_{\hat{e}_i}(y) \right|\end{aligned}$$

First, notice

$$\frac{w_e}{p_e} Q_e(L_G^{-1/2}x) = \frac{w_e}{p_e} \max_{u,v\in\binom{e}{2}} \langle L_G^{-1/2}x, b_{u,v}\rangle^2 = \max_{u,v\in\binom{e}{2}} \langle x, z_{u,v}^e\rangle^2$$

where $z_{u,v} = L_G^{-1/2} b_{u,v}$ and $z_{u,v}^e = \sqrt{w_e/p_e} z_{u,v}$. If we let

$$N_k(x) = \max_{u,v\in e_k} |\langle x, z_{u,v}^{e_k}\rangle|$$

Then,

$$Q_{\tilde{H}}(L_G^{-1/2}x) = \frac{1}{m}\sum_{i=1}^k N_i(x)^2.$$

Similarly, the object we need to study is

$$\frac{1}{m}\sum_{i=1}^k s_i N_i(x)^2$$

for Radamacher random variables $\{s_i\}_{1\leq i\leq m}$. So, we just need to the expected value of this quantity over the space $T = \{x : Q_H(L_G^{-1/2}x) \leq 1\}$. This is where the chaining is used. But the details is beyond the scope of this course.