What is this class about?

Concentration of random matrices.

Scalar valued vector valued matrix valued	because the concentration is of a statistic that's sensible when viewing the object as a matrix.

Spectrum, λ_1 (largest eigenvalue), spectral norm, ...

Our focus is on applications to theoretical computer science

(1) algorithm design: randomized numerical lin algebra, graph sparsification, graph partitioning, analysis of algo on rand inputs, ...

(2) statistical estimation: estimation from samples covariance/moments, sparse regression, compressive sensing, ...

(3) coding theory/combinatorics: girth problems, ...

④ Complexity Theory: construction/existence of expander graphs,

⑤ Quantum Information: quantum query algos.

⑥ discrepancy theory: beating the union bound

Plan at the begining:

Basic Matrix Conc Inequalities, Proofs, Tight cases

Let's start by recalling three standard, simple concentration inequalities for which we will soon write the matrix analogs

Khintchine inequality "Simplest"?

$a_1 \ldots a_n$ real numbers.

$g_1 \ldots g_n$ independent Gaussian random variables.

$b_1 \ldots b_n$ independent, uniform ± 1 rand vars

"Rademacher"

Then,

$$\mathbb{E}\left[\left|\sum_i g_i a_i\right|\right] = \sqrt{\sum_i a_i^2} \cdot \sqrt{\frac{2}{\pi}}$$

$$c \cdot \sqrt{\sum_i a_i^2} \leq \mathbb{E}\left[\left|\sum_i b_i a_i\right|\right] \leq C \cdot \sqrt{\sum_i a_i^2}$$

$$c = \frac{1}{\sqrt{2}}, \quad C = 1.$$

Bernstein Inequality

more general than Khintchine

$X_1, \ldots, X_n$ independent, zero-mean, $X_i \le L$ almost surely.

$$\sigma \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^n \mathbb{E} X_i^2}$$

Then,
$$\Pr\left[\sum_{i=1}^n X_i \geqslant 2t\cdot\sigma\right] \le e^{-t^2}$$

for all $0 \le t \le \frac{\sigma}{2L}$.

Chernoff Bound

natural for Bernoulli
deviation relative to the mean

$X_1, \ldots, X_n$ indep, $X_i = \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{o/w} \end{cases}$

$$\mu = \mathbb{E}[\Sigma_i X_i] = \Sigma_i p_i$$

Then,

$$\Pr[\Sigma_i X_i \geq (1+\delta)\mu] \leq e^{-\frac{\mu \cdot \delta^2}{2+\delta}} \quad \forall \delta > 0$$

$$\Pr[\Sigma_i X_i \leq (1-\delta)\mu] \leq e^{-\mu \cdot \delta^2/2} \quad \forall 0 < \delta < 1$$

We first start with the matrix analog of Khintchine inequality. We will later use it to establish matrix analogs of Bernstein and Chernoff too.

Matrix Khintchine Inequality

[Lust-Piquard, Pisier '91]

There are absolute constants $c, C > 0$ s.t. the following holds:

For any $A_1, \ldots, A_n$, $m \times m$ real symmetric matrices

$$c \cdot \sigma(\vec{A}) \leq \mathbb{E}_{\vec{g}} \left\| \sum_{i=1}^{n} g_i A_i \right\|_2 \leq C \sqrt{\log m} \cdot \sigma(\vec{A})$$

Here $\vec{g}$ is a vector of standard Gaussian random variables. $\sigma^2(\vec{A}) = \max\{\|\sum_i A_i A_i^T\|_2, \|\sum_i A_i^T A_i\|_2\}$

Concrete, we can take $c = \frac{1}{\sqrt{2}}, C = \sqrt{2e}$

Further, there exist $c', C' > 0$ s.t.

$$c \cdot \sigma(\vec{A}) \leq \mathbb{E}_{\vec{b}} \|\sum_i b_i A_i\|_2 \leq C \cdot \sqrt{\log m} \cdot \sigma(\vec{A})$$

where b_i are uniform, indep $\{\pm 1\}$-rand. vars.

Remark: (the $\sqrt{\log m}$ dependence is tight)

Ex 1: $A_i = e_i e_i^T$ where $e_i = (0, \ldots, 1, \ldots 0)$ (i^{th} pos)

Then $\sum_i g_i A_i$ is a matrix of indep standard Gaussians on the diagonal

$$\mathbb{E} \|\sum_i g_i A_i\|_2 = \mathbb{E}_g \max\{|g_1|, \ldots, |g_m|\}$$

$$\gtrsim \sqrt{\log m}$$

Ex 2: (lower tail is tight).

$$A_{i,j} = e_i e_j^T \quad \forall\ i \neq j.$$

Then $\mathbb{E} \| \sum_i g_{i,j} \cdot A_{i,j} \|_2 = (2+o(1))\sqrt{m}$

while $\sigma(\vec{A}) = \sqrt{m-1}$

Ex 3: (tightness in the Rademacher case)

Ⓐ Suppose $m = 2^n$.

Choose $V_1, \ldots, V_m \in \{\pm 1\}^n$ to be the set of all possible $\{\pm 1\}$ entry vectors.

set $A_i(r,s) = V_r(s)$.

Then $\| \sum_i b_i \cdot A_i \|_2 = \max_{1 \le j \le m} \langle b, V_j \rangle$

$= n \quad \forall\, b.$

$= \| \sum_i A_i^2 \|_2^{1/2} \cdot \sqrt{\log m}.$

(b) Suppose $m << 2^n$.

Then,

Application to Graph Sparsification

Definition (ε-cut sparsifier)

Let $G(V,E)$ be an undirected, unweighted graph on $n = |V|$ vertices.

A weighted graph $G'(V, E', w: E' \to \mathbb{R}_+)$ is an ε-cut sparsifier for G if $\forall S \subseteq V$

$$(1-\varepsilon)|E_G(S,\bar{S})| \le |E_{G'}(S,\bar{S})| \le (1+\varepsilon) \cdot |E_G(S,\bar{S})|$$

Here, $E_G(S,\bar{S}) = \{e = \{i,j\} \in E \mid i \in S, j \in \bar{S}\}$

& $|E_{G'}(S,\bar{S})| = \sum_{e \in E' \cap E_G(S,\bar{S})} w_e$

Remark: ε-cut sparsifiers satisfy a relative error guarantee $\to$ small cuts have a smaller absolute error. This is a key difficulty in constructing them.

Theorem: For every (weighted) graph $G(V,E)$ there is an ε-cut sparsifier with $m = O\left(\frac{n}{\varepsilon^2}\log(n)\right)$ edges. $O\left(\frac{n}{\varepsilon^2}\right)$ known!

History:

1993,95	combinatorial methods + sampling	Karger, Karger-Stein
2006	spectral method + sampling	Spielman Srivastava
2008	convex programming	Batson-Spielman Srivastava

Today: an iterative halving argument

Let's present the key piece first

Towards this let's first relate cut sparsification to quadratic form approximation.

A_G: adjacency matrix, $n \times n$

$$A_G(i,j) = \begin{cases} w_e & \text{if } e=\{i,j\}\in E \\ 0 & \text{o/w} \end{cases}$$

$$L_G = D_G - A_G$$

where $\underbrace{D_G(i,i)}_{\text{diagonal matrix of weighted degrees}} = \sum_{e:e\ni i} w_e$, $D_G(i,j)=0$ if $i\neq j$

For edge i,j, let $\underbrace{L_e}_{\text{edge Laplacian}} = \begin{array}{c} \\ i \\ j \\ \\ \end{array}\begin{array}{c} \\ \\ \\ \\ \end{array}\!\!\begin{array}{cccc} & i & j & \\ \end{array}$

$$L_e = \begin{array}{c|cccc} & & i & j & \\ \hline & 0 & & & 0 \\ i & & 1 & -1 & \\ j & & -1 & 1 & \\ & 0 & & & 0 \end{array}$$

<u>Obs^n</u>: $x^T L_e x = (x_i - x_j)^2$

<u>Obs^n</u>: L_e is a rank 1, positive semidefinite

<u>Obs^n</u>: $L_G = \sum_{e\in E} L_e$, is also PSD.

Obs^n: If G is connected, L_G has a kernel of dim 1.

Obs^n (Cuts vs ± 1-vector quad forms)

Let $S \subseteq V$. Define $x = \mathbb{1}(S) - \mathbb{1}(\bar{S})$

$$x_i = \begin{cases} 1 & \text{if } i \in S \\ -1 & \text{o/w} \end{cases}$$

Then $$x^T L_{G'} x = \sum_{\{i,j\}=e \in E} w_e \cdot (x_i - x_j)^2$$

$$= 4 \cdot |E_{G'}(S, \bar{S})|$$

Def (ε-Spectral Sparsifier)

$G'(V, E')$ is an ε-Spectral sparsifier of $G(V, E)$ if $(1-\varepsilon) L_G \preccurlyeq L_{G'} \preccurlyeq (1+\varepsilon) \cdot L_G$

$$\Leftrightarrow \forall z, \; z^T L_{G'} z \in (1 \pm \varepsilon) \cdot z^T L_G z$$

Spectral sparsifier $\equiv$ approximate <u>all</u> quad forms

Cut sparsifier $\equiv$ approximate ± 1-vector quad forms

Splitting Lemma: Let $G(V, E, w)$ be a weighted graph with $m \geq$ edges. Then there's a weighted graph $G'(V, E', w')$ such that

① $|E'| \leq \left(\frac{7}{8}\right) \cdot |E|$.

② $$(1-\delta) L_G \preccurlyeq L_{G'} \preccurlyeq (1+\delta) \cdot L_G$$

for $\delta = O(\sqrt{\log n}) \cdot \sqrt{\frac{n}{m}}$

Let's first see how to complete the proof using this lemma.

Sparsification Algo:

1) Set $E_0 = E$. Let $m_0 := |E|$
$W^0 = \vec{\mathbb{1}}$.

2) For $i = 1, 2, \ldots$

(a) while $m_{i-1} \geq \frac{c \cdot n \cdot \log n}{\varepsilon^2}$

apply splitting lemma on
$G(V, E_{i-1}, W^{i-1})$

to obtain E', W'

(b) set $E_i = E'$, $W^i = W'$

3) Go to step 1.

Let i^* be the final iteration.

Analysis:

Condition on $\frac{7}{8} m_{i-1} \geq m_i \geq \frac{5}{8} m_{i-1}$

$\forall\, i \leq i^* \to$ event occurs with $1 - 1/n$ prob

enough to ~~bound~~

$$\Big\| \sum_{e \in E_{i^*}} w_e^{i^*} \cdot \tilde{L}_e - \Pi_1 \Big\|_2$$

$$\leq \sum_{i=1}^{i^*} \Big\| \sum_{e \in E_i} w_e^{i} \cdot \tilde{L}_e - \sum_{e \in E_{i-1}} w_e^{i-1} \cdot \tilde{L}_e \Big\|_2$$

$$\lesssim \sum_{i=1}^{i^*} \Big\| \sum_{e \in E_{i-1}} w_e^{i-1} \cdot \tilde{L}_e \Big\| \cdot \underbrace{C'}_{\text{abs constant}} \sqrt{\log n^{-1}} \sqrt{\frac{n}{m_{i-1}}^{-1}} \quad (1)$$

For large enough c s.t. $m \geq \frac{c \cdot \log n \cdot n}{\varepsilon^2}$

Let's argue by induction on t

$$\sum_{i=1}^{t} \left\| \sum_{e \in E_{i-1}} w_e^{i-1} \cdot \tilde{L}_e \right\|_2 \cdot c' \sqrt{\log n} \sqrt{\frac{n}{m_i}}$$

$$\leq c'' \cdot \sqrt{\log n} \cdot \sqrt{\frac{n}{m_t}} \quad \text{— (1)}$$

True easily for $t=1$.

Let's argue the inductive step.

If true for $t-1$, then $\forall i \leq t-1$

$$\left\| \sum_{e \in E_i} w_e^i \cdot \tilde{L}_e \right\|_2 \leq \left(1 + c'' \sqrt{\log n} \sqrt{\frac{n}{m_i}}\right)$$

≤ 2 for large enough c.

<u>So</u>: LHS of (1)

$$\leq 2 \cdot \sum_{i=1}^{t} c' \cdot \sqrt{\log n} \cdot \sqrt{\frac{n}{m_{i-1}}} \leq 2 O\left(\frac{\sqrt{n \log n}}{\sqrt{m_t}}\right)$$

as m_i decrease geometrically. ☺

$$= C^{4} \cdot \sqrt{\log n^{I}} \cdot \sqrt{\frac{\Lambda}{m_t}}$$ as derived.

Let's now prove the splitting lemma.

Proof: WLOG assume G is connected.

1. From relative to additive error.

Def (normalized edge Laplacian)

$$\tilde{L}_e = L_G^{+/2} \cdot L_e \cdot L_G^{+/2}$$

Aside: $L_G = \sum \lambda_i \cdot v_i v_i^T$ be EVD

then $L_G^{+} = \sum_{i:\lambda_i>0} \lambda_i^{-1} \cdot v_i v_i^T$ "pseudo inverse"

$L_G^{+/2} = \sum_{i:\lambda_i>0} \lambda_i^{-1/2} \cdot v_i v_i^T$ "pseudo inverse square root"

Obs^n: $\sum_{e\in E} w_e \cdot \tilde{L}_e = L_G^{+/2} \cdot L_G \cdot L_G^{+/2}$

$= \sum_{i:\lambda_i>0} v_i v_i^T \overset{\text{def}}{=} \mathbb{I}_\perp$

Obs^n: Let $E' \subseteq E$.

Then $\sum_{e\in E'} w_e' \cdot L_e \overset{1\pm\delta}{\approx} \sum_{e\in E} w_e \cdot L_e$

iff $\left\| \sum_{e\in E'} w_e' \cdot \tilde{L}_e - \sum_{e\in E} w_e \cdot \tilde{L}_e \right\|_2 \leq 2\delta$

Proof of obs^n: Note $L_G \vec{\mathbb{1}} = L_{G'} \vec{\mathbb{1}} = 0$ & L_G has rank $n-1$.

So suppose $x \in \mathbb{R}^n$, $x \perp \vec{\mathbb{1}}$.

Then

$$\forall x: \left| x^T \Big(\sum_{e \in E'} w_e' L_e \Big) x - x^T \Big(\sum_{e \in E} w_e L_e \Big) x \right| \leq \delta \cdot x^T \Big(\sum_{e \in E} w_e L_e \Big) x$$

$$\Leftrightarrow \underline{\forall x}: \left| x \cdot L_G^{1/2} \cdot \Big(\sum_{e \in E} w_e' \cdot \tilde{L}_e \Big) \cdot L_G^{1/2} x - x^T \cdot L_G^{1/2} \cdot \Big(\sum_{e \in E} w_e \tilde{L}_e \Big) L_G^{1/2} \cdot x \right| \leq \delta \cdot (L_G^{1/2} x)^T (L_G^{1/2} \cdot x)$$

$$\Leftrightarrow \forall y: \left| y^T \Big(\sum_{e \in E} w_e' \cdot \tilde{L}_e \Big) y - y^T \Big(\sum_{e \in E} w_e \cdot \tilde{L}_e \Big) y \right| \leq \delta \cdot \|y\|_2^2$$

$$\Leftrightarrow \left\| \sum_{e \in E'} w_e' \tilde{L}_e - \sum_{e \in E} w_e \cdot \tilde{L}_e \right\|_2 \leq \delta.$$

Step 2: "subsample less important edges"

Def: (importance) Given $G(V,E)$, $w: E \to \mathbb{R}_+$

$$I(e) \stackrel{\text{def}}{=} \| w_e \cdot \tilde{L}_e \|_2$$

Claim: $\sum_{e \in E} I(e) =$

Pf: $\sum I(e) = \sum_{e \in E} \| w_e \cdot \tilde{L}_e \|_2$

$$= \sum_{e \in E} \mathrm{Tr}(w_e \cdot \tilde{L}_e)$$

$$= \mathrm{Tr}\left(\sum_{e \in E} w_e \cdot \tilde{L}_e\right)$$

$$= \mathrm{Tr}\left(L_G^{+/2} \cdot L_G L_G^{+/2}\right) = \mathrm{Tr}(\mathbb{I}_1)$$

$$= n - 1.$$

Corollary: For at least $\frac{m}{2}$ edges,

$$I(e) \leq 2 \cdot \frac{n-1}{m}.$$

Pf: Markov + claim.

Sampling Scheme: 1) Let $E^{good} = E \cap \{e : I(e) \leq \frac{2(n-1)}{m}\}$

2) Let $s \in \{\pm 1\}^{E^{good}}$ be chosen uniformly at random.

3) Set $E' = E \setminus E^{good} \cup \{e \in E^{good} \mid s_e = 1\}$

4) $$w_e' = \begin{cases} w_e & \text{if } e \in E \setminus E^{good} \\ 2w_e & \text{if } e \in E' \cap E^{good} \\ 0 & \text{o/w.} \end{cases}$$

<u>Claim</u>: If $m \geq n \log n$, then with probability at least $1 - 1/n^{100}$,

$$\frac{3}{4}m \leq |E'| \leq \frac{7m}{8}$$

<u>Pf</u>: Chernoff bound.

<u>Step 3</u>: Error Analysis

$$\left\| \sum_{e \in E'} w_e' \cdot \tilde{L}_e - \sum_{e \in E} w_e \cdot \tilde{L}_e \right\|_2$$

$$= \left\| \sum_{e \in E^{good}} \delta_e w_e \tilde{L}_e \right\|_2$$

NCK

$$\lesssim \sqrt{\log n} \left\| \sum_{e \in E^{good}} (w_e \cdot \tilde{L}_e)^2 \right\|_2^{1/2} \quad (1)$$

Now, since $\forall e \in E^{good}$, $w_e \cdot \tilde{L}_e \prec I \otimes II$

$$\sum_{e \in E^{good}} (w_e \cdot \tilde{L}_e)^2$$

$$\preceq I(e) \cdot \Pi \cdot \sum_{e \in E^{good}} w_e \cdot \tilde{L}_e$$

$$\preceq I(e) \sum_{e \in E} w_e \cdot \tilde{L}_e = I(e) \cdot \Pi_\perp$$

Thus, $\left\| \sum_{e \in E^{good}} (w_e \cdot \tilde{L}_e)^2 \right\|_2$

$$\leq 2 \cdot \left(\frac{n-1}{m}\right).$$

Thus, by (1),

$$\left\| \sum_{e \in E'} w_e' \cdot \tilde{L}_e - \sum_{e \in E} w_e \cdot \tilde{L}_e \right\|_2$$

$$\preceq O\left(\sqrt{\log n}\right) \cdot \sqrt{\frac{n-1}{m}}.$$

$\square$

Notes:

The algorithm suggests a natural way to improve C & the key reason we choose to present this version instead of the Spielman-Srivastava original).

→ instead of choosing S randomly in the splitting step, we could try to find the best S that drops m by a constant factor while controlling the spectral norm of the signing.

This is one way to view the idea of Batson-Spielman-Srivastava that yields a "twice ramanujan sparsifier" with $O\left(\frac{n}{\varepsilon^2}\right)$ edges.