(Lec 14: Sum-of-Squares vs Low-deg Polynomials)

Last time: Spectral Algorithms

average-case problem : dist^n over instances

$D_{planted}$: (YES) instances

D_{null} : (NO) instance → refute the existence of a sol^n

find the planted solution

given an instance $G \sim D_{planted}$ or $G \sim D_{null}$ → report correctly the generating dist^n

$\underbrace{\text{distinguishing}}$ $\lesssim$ $\left\{ \begin{array}{l} \underline{\text{planted}} \\ \underbrace{\text{refutation}} \end{array} \right.$

Last time: Spectral algorithms

Today: $\underbrace{\text{"Low-degree polynomial"}}$ → distinguishing

---

Clique in random graphs

D_{null}: $G(n, \frac{1}{2})$

$D_{planted}$: $G(n, \frac{1}{2}) + \omega$-clique

random 3-XOR

$n \to$ # vars

$m \to$ # constraints

which triples are included?

m-edge 3-uniform hypergraph

What are RHS for each constraint

Today H is fixed & same for both Dplanted & Dnull

$c_1, c_2, \ldots, c_m \in \binom{[n]}{3}$

$b_1, b_2, \ldots, b_m$

Think of $m \geq \left(\frac{Cn}{\varepsilon^2}\right)$.

D_{null}: Choose $b_i \in \{\pm 1\}$ uniform & indep

$D_{planted}$: $x^* \in \{\pm 1\}^n$ uniform

for each c_i, $b_i = x^*_{c_i}$

Low-degree Polynomial

$$\mathcal{A}(G) \in \{0, 1\}$$

"null" "planted"

Correctness:
$$\Pr_{G \sim D_{null}} [\mathcal{A}(G) = 0] \geq 1 - o_n(1).$$

$$\Pr_{G \sim D_{planted}} [\mathcal{A}(G) = 1] \geq 1 - o_n(1)$$

degree d polynomial : $\mathcal{A}(G)$ must be a degree d "polynomial in G"

algo: $\qquad$ a polynomial

clique: $G \sim \{0,1\}^{\binom{n}{2}}$: $A(G)$ can compute deg d polynomials over $\{0,1\}^{\binom{n}{2}}$

3-XOR: $\bar{b} \in \{\pm 1\}^m$: $A(\bar{b})$ can compute a deg d polynomial over $\{\pm 1\}^m$.

how powerful are such algorithms?

$$\mathbb{1}\left[\#\text{edges}(G) \gtrsim \text{threshold}\right]$$

$$G \sim G(n, \tfrac{1}{2}) : \underbrace{\Big[\overset{\text{mean}}{\tfrac{1}{2}\tbinom{n}{2}} + O(\sqrt{\log n}) \cdot n\Big]}\quad \text{w.p} \geq 1 - \tfrac{1}{n}.$$

for what ω is #edges in $\underbrace{G \sim G(n, \tfrac{1}{2}) + \omega\text{-clique}}_{\frac{1}{2}\binom{n}{2} + \binom{\omega}{2} \,\sim\, \omega^2}$

$$\omega^2 \gg n\sqrt{\log n}$$

$$\sim \omega \gg \sqrt{n} \cdot \log(n)^{1/4}.$$

linear polynomial

$$\mathbb{1}\left(\text{max-degree} \geq \text{threshold}\right)$$

Is max-degree(G) a low-deg poly function.

$\ell_p \Leftrightarrow \ell_\infty$ trick

max-degree: $\left\| \begin{pmatrix} d_1 \\ | \\ | \\ d_n \end{pmatrix} \right\|_\infty \to$ vector of degrees

$f_t(G) = \sum_i \left(d_i - \frac{n}{2}\right)^t \leftarrow$ polynomial of deg t in G.

__Obs__^{n}: $n \cdot d_{max}^t \geq f_t(G) \geq d_{max}^t$

__Suppose__: w.p. $1-o(1)$, $d_{max}(G) - \frac{n}{2} \leq B\sqrt{n}$

over $G \sim G(n, \frac{1}{2})$ $\qquad B = O(\sqrt{\log n})$

w.p. 1 $\quad d_{max}(G) - \frac{n}{2} \geq \omega$

$G \sim G(n, \frac{1}{2})$

$+\ \omega$-clique

$\omega \not\geq 2B \cdot \sqrt{n}$

null: $f_t(G) \sim \sum_i \left(d_i - \frac{n}{2}\right)^t \leq n\, \underbrace{B^t (\sqrt{n})^t}$

planted: $f_t(G) = \sum_i (d_i - \frac{n}{2})^t \geq \omega^t \geq \underbrace{(2B\sqrt{n})^t}_{= 2^t \cdot B^t \cdot \sqrt{n}^t}$

$$2^t \cdot B^t \cdot (\sqrt{n})^t > n \cdot B^t \cdot (\sqrt{n})^t$$

$$2^t \gg n \quad \text{or} \quad t \gg \log(n).$$

<u>Conclusion:</u> instead of $d_{max} - \frac{n}{2}$

$\rightarrow$ consider $f_t(G)$ for $t \sim \log(n)$

<u>Question:</u> $\log(n)$-deg poly can simulate "max-deg". Are they a "efficient"?

How long does it take to compute a deg t polynomial in G? $\longrightarrow n^{O(t)}$

(we assumed that poly is described as a vector of its coefficients)

there are implicit descriptions that can allow you to compute polynomial ... $\sim 2^{O(t)} \text{poly}(n)$

$\rightarrow$ we'll think of degree $O(\log(n))$ polynomials as "efficient" algorithms.

<u>Last time</u>: $\lambda_{max}(A(G))$

We proved $\lambda_{max}(G) \leq O(\sqrt{n})$ w.p. $1-o(1)$

<u>over $G \sim G(n, \frac{1}{2})$</u>

$\rightarrow$ if $\omega \gg C \cdot \sqrt{n}$ for $C > 0$ large enough

then $\lambda_{max}(G)$ is a distinguisher..

<u>Fact</u>: A is symm $n \times n$ matrix, $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ are eigenvalues

then eigs of A^{2k} are: $\underline{\lambda_1^{2k}} \geq \underline{\cdots \geq \lambda_n^{2k}}$

<u>Then</u> (as a consequence): $\sum_i \lambda_i^{2k}$ is a degree

$2k$ polynomial in entries of A.

$2k$ polyn

Pf: $\operatorname{tr}(A^{2k}) = \sum_i \lambda_i^{2k}$

is a deg $2k$ polynomial in entries of A

$G \sim G(n, \frac{1}{2})$: $\operatorname{tr}(A^{2k})$

$$= \sum_i \lambda_i^{2k} \le (C \cdot \sqrt{n})^{2k} \cdot n$$

$G \sim$ planted: $\operatorname{tr}(A^{2k}) = \sum_i \lambda_i^{2k}$

$$> \omega^{2k}.$$

$\underline{\omega > 2 \cdot C \cdot \sqrt{n}}$: if $2^t \gg n$

or $t \gg \log(n)$

then $\lambda_{max}(G)$ is a distinguisher

Comment: We rely on rigidity of $\lambda_{max}(A)$ for Wigner A to actually conclude that this distinguisher works for $\omega \gg C \cdot \sqrt{n}$

$(1-o(1))$

$C \to$ abs constant.

Conclusion: For 4-XOR we saw a $\lambda_{max}(A)$ distinguisher: $A \in \mathbb{R}^{n^2 \times n^2}$ matrix built from input instance $\rightarrow$ can again be approximated well by a $O(\log(n))$ deg polynomial.

Spectral distinguishers: "no easy way to find the best"

Low deg polynomials: given $d \in \mathbb{N}$, can explicitly describe the best distinguisher

Theorem: Let $D_{planted}$ & D_{null} be dist^n's on $\{\pm 1\}^n$. Then we can explicitly describe the sol^n to the following problem.

$$\boxed{\max_{p:\ p \text{ is a deg } d \text{ poly}} \ \mathbb{E}_{G \sim D_{planted}} p(G)}$$

Computes the best deg d poly distinguisher s.t. $\mathbb{E}_{G \sim D_{null}} p(G) = 1.$

$\mathbb{E}_{G \sim D_{null}} (p(G) - 1)^2 = 1$

Pf: Suppose D_{null} is uniform on $\{\pm 1\}^n$.

then recall the fourier basis for D_{null}.

$\forall S \subseteq [n], \quad \chi_S, \quad \mathbb{E}_{D_{null}} \chi_S = 0$

$\mathbb{E}_{D_{null}} \chi_S^2 = 1$

$\mathbb{E}_{D_{null}} \chi_S \chi_T = \begin{cases} 0 & \text{if } S \neq T \\ 1 & \text{o/w.} \end{cases}$

$$p(G) = \sum_{|S| \leq d} \hat{p}(S) \cdot \chi_S$$

$$\mathbb{E}_{D_{null}} p(G) = 1 \iff \hat{p}(\phi) = 1$$

$$\mathbb{E}_{D_{planted}} p(G) = \sum_{|S| \leq d} \hat{p}(S) \underbrace{\mathbb{E}_{D_{planted}} \chi_S}$$

$1 \quad (= p_\phi) \quad 1$

$$\left\langle \begin{pmatrix} p_{\{i\}} \\ p_{\{ij\}} \\ \vdots \end{pmatrix}, \begin{pmatrix} S \mapsto \mathbb{E}_{D_{planted}} x_S \end{pmatrix} \right\rangle$$

$$\rightarrow \underbrace{p \propto \left(\mathbb{E}_{D_{planted}} X_S \right)}_{p \,:\propto\, \boxed{D^{\le d}_{planted}}}$$

$\rightarrow$ "degree d projection of $D_{planted}$"

$$\underline{D_{planted}} = \underline{\sum \hat{D}_{planted}(S) \cdot X_S}$$

$$\hat{D}_{planted}(S) = \mathbb{E}_{x \sim \{\pm 1\}^n} D_{planted}(x) \cdot X_S$$

$$= \frac{1}{2^n} \cdot \sum_x D_{planted}(x) \cdot X_S$$

$$= \sum_x \left(\frac{D_{planted}}{2^n} \right) \cdot X_S$$

$$= \frac{1}{2^n} \cdot \mathbb{E}_{D_{planted}} X_S$$

For arbitrary D_{null} ← use the right orthogonal polynomial basis

<u>Sol^n</u>: $\boxed{\text{truncated relative planted density}}$

<u>Conclusion:</u> deg d polynomials distinguish bet^n $D_{planted}$ & D_{null} iff

$$\underbrace{\left\| D_{planted}^{\le d} \right\|^2}_{\text{optimal deg } d \text{ polynomial}} = \sum_{d \ge |S| \ge 1} \left(\mathbb{E}_{D_{planted}} X_S \right)^2 \gg 1$$

$$\| $$

$$\underbrace{\mathbb{E}_{D_{planted}} f} \gg \underbrace{\mathbb{E}_{D_{null}} f}_{= 1} + B \underbrace{\sqrt{\mathbb{E}_{D_{null}} (f-1)^2}}_{\downarrow\, 1}$$

Is there a connection bet^n low deg poly & SoS?

Yes, probably:

Yes, probably

"Pseudocalibration"

<u>Historical Note</u>: <u>lower bounds for high deg</u>

clique on $G(n,\frac{1}{2})$.

Sparse PCA

Tensor PCA

<u>Issues</u>: <u>natural constructions of pseudo-dist^{n}s failed</u>...

"low degree method"

vs

SoS lower bounds

→ Suppose ∃ a Dplanted s.t. degree $O(d \cdot \log(n))$ polynomials <u>fail</u> to distinguish bet^n Dplanted & Dnull. Then there's a "natural" candidate p.d. of degree d that's consistent with constraints for Dnull.

Sketch for 3XOR

Pseudocalibration Conjecture: for a large class of avg-case problems $\to D_{null}$, if there's a $D_{planted}$ s.t. $D_{planted} \overset{d \log n}{\sim} D_{null}$ then there's a degree d SoS lower ~~bound~~ for refuting instances from D_{null}.

Pseudocalibration for 3XOR

$$\boxed{m \geq \frac{Cn}{\varepsilon^2}}$$

$$\mathcal{A} : \begin{cases} b_i x_{C_i} = 1 & \forall\ 1 \leq i \leq m \\ x_i^2 = 1 & \forall\ 1 \leq i \leq n \end{cases}$$

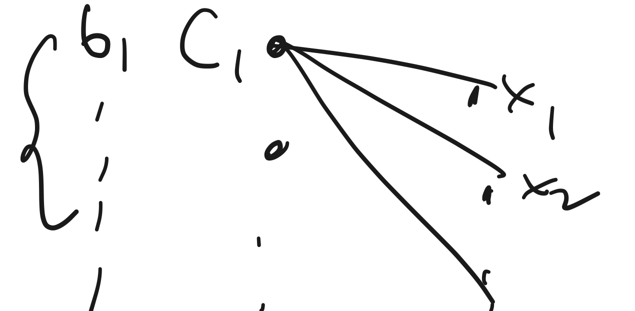
$b_i \sim D_{null}$

In earlier lecture, we proved that $\Omega(n)$ deg SoS cannot refute $\mathcal{A}$ so long as $\mathcal{H}$ is "expanding".

Expansion

$\forall\ T \subseteq [m]$, $|T| \leq \frac{n}{10^4}$, T

$|\bigcup C_i| \geq 1.7|T|$



$$\Big|\bigcup_{i\in T} C_i\Big| \geq 1+|T|$$

$\cdot x$

$b_m C_m$

<u>Pseudocalibration</u>: use Dplanted to help come up with a pseudo-distⁿ sat $\mathcal{A}$.

<u>Idea</u>: Set $\tilde{\mathbb{E}}_{\mu} X_S$ = "low-degree truncations of the planted distⁿ."

$\forall \underline{|S|\leq d}$

Fix $S\subseteq [n]$, $|S|\leq d$.

$$\tilde{\mathbb{E}}_{\mu} X_S[b] = \sum_{T\subseteq[m]} \hat{f}(T)\cdot \bar{b}_T$$

$\prod_{i\in T} b_i$

$$\widehat{\mathbb{E}_{\text{Dplanted}} X_S[b]}(T)$$

$$= \mathbb{E}_{(x^*,b^*)\sim \text{Dplanted}} \cdot X^*_S \cdot b^*_T$$

$$\mathbb{E}\, X_S[b] = \sum \Big(\mathbb{E} \cdot X^*_S \cdot b^*_T\Big)\cdot b_T.$$

$\mathcal{D}_{planted}$ $\qquad T \in [m]$ $\qquad \mathcal{D}_{planted}$

I ~ planted

$x^* \in \{\pm 1\}^n$ uniformly $\quad \Big| \quad C_1 \text{ ——— } C_m$

$b_i^* = x^*_{C_i} \quad \Big| \quad \underline{b_1} \qquad \underline{b_m}$

For any $\mathcal{I}$: $\quad \underline{\underline{x_S}} = \underline{x_S^*} = \prod_{i \in S} x_i^*$

$$\underbrace{\mathbb{E}_{\mathcal{D}_{planted} | b}}\, x_S = x_S^* \cdot \longleftarrow$$

$$\tilde{\mathbb{E}}_{\mu} x_S [b] \overset{def}{=} \sum_{|T| \le \tau} \Big(\mathbb{E}_{\mathcal{D}_{planted}} x_S^* \cdot b_T^* \Big) \cdot b_T$$

pseudo-distr being defined

truncation threshold.

"μ pretends as if the instance I was chosen according to the planted distn"

One consequence of this def.

$\forall\ f(x, \bar{b})$. polynomials such that
$\deg(x) \le d$
$\deg(\bar{b}) \le \tau$.

$$\mathop{\mathbb{E}}_{\bar{b} \sim D_{null}} \mathop{\tilde{\mathbb{E}}}_{\mu} f(x, \bar{b}) = \mathop{\mathbb{E}}_{(x,\bar{b})\, D_{planted}} f(x, \bar{b}).$$

$$\tilde{\mathbb{E}}\, x_S(b) = \sum_{|T| \le \tau} \left(\mathop{\mathbb{E}}_{D_{planted}} x_S^* \cdot b_T^* \right) b_T$$

$$\mathop{\mathbb{E}}_{D_{planted}} x_S^* \cdot b_T^* = \mathop{\mathbb{E}}_{D_{planted}} x_S^* \cdot \prod_{i \in T} b_i^*$$

$D_{planted}$ $\quad$ $D_{planted}$

$$= \mathbb{E}_{D_{planted}} \underbrace{X_S^*; \prod_{i \in T} x^*_{C_i}}$$

non-zero iff every variable appears with an even power.

$\Leftrightarrow$ Variables that appear odd # of times in $\bigcup_{i \in T} C_i$ are exactly equal to S.

$$\Leftrightarrow \underset{i \in T}{\Delta} C_i = S$$

$\hookrightarrow$ $X_{C_i} = b_i$ for $i \in T$ "derive" a value for X_S.

Claim: If $\tau \ll \frac{n}{4}$ & H is expanding

$\frac{}{10^{?}}$

then $\tilde{\mathbb{E}}_{\mu} X_S = \begin{cases} b_T & \text{for the } \underline{\text{unique}} \text{ subset } T \text{ of constraints that derive } X_S \\ 0 & \text{o/w} \end{cases}$

$\downarrow$

<u>pseudocalibrated pseudo-dist^n</u>

$\equiv$ <u>Grigoriev's construction</u>

Summary: 1) Low-deg poly method is surprisingly powerful

2) Can be understood relatively easily

3) seems to relate to the performance of more powerful algos like SoS.

$$\tilde{\mathbb{E}}_{\mu} x_S[b] = \sum_{|T| \le \tau} \widehat{\tilde{\mathbb{E}}_{\mu} \cdot x_S} \; b_T \, .$$

pseudo-expectation "hat".

Average-Case Complexity

* $\boxed{\text{Pseudocalibration Conjecture}}$

SQ model
Belief Propagation / Local algos
SoS SDPs
Lowdeg Method

Worst-case 3XOR: $\Leftarrow$

* SQ $\sim$ Low degree method

* Community Detection

Worst-Case Complexity.

[illegible]

UGC? deg 4 SOS?

S.S.E.

Sparsest Cut/ Expansion : ARV $O(\sqrt{\log n})$
deg 4 SoS

Max-Cut : Is there a $2^{n^{1-\delta}}$ algo for beating GW for Max-Cut?

U.G. $\Leftarrow$ $2^{n^{\varepsilon}}$ algo for solving UG using G.C. rounding.

Is UG strictly easier than Max-Cut?

We know UG $\lesssim$ Max-Cut

Pseudocalibration Conjecture : for a large class of any class.

if degree d SoS can distinguish betn Dplanted & Dnull then $\exists$

$n^{O(d)} \times n^{O(d)}$ matrix A with entries that are degree d polynomial s.t.

$\lambda_{max}(A)$ is a distinguisher

$\downarrow$

try to apply our trick $\ell_p \Leftrightarrow \ell_\infty$ from today

Gap betn $\lambda_{max}(A)$ & $\|A\|_2$ is the main issue here

---

$D_{planted}$: (x^*, I).

$\downarrow$

"well behaved under random perturbations"

$G \sim G(n, \frac{1}{2}) + \omega$-clique

$\downarrow$

$S \subseteq [n]$, $|S| \le n^{0.99}$

and re-randomize edges that do not go within S. $\to$ then graph still has a large planted clique.

$f: \{\pm 1\}^n \to \mathbb{R}$ ↗ Ryan's Book!

$$T_\rho f(x) = \boxed{\mathbb{E}_{y \sim N_\rho(x)} f(y)}$$

flip every coordinate of x indep w.p. $1-\rho$

$$\mathbb{E}_{x \sim \{\pm 1\}^n} T_\rho f(x) \cdot f(x) \le \sqrt{\mathbb{E}\, T_\rho f} \sqrt{\mathbb{E} f^2}$$

f is noise-stable if & only if

$\exists\, g$: low deg polynomial st

$\mathbb{E}(f-g)^2 < \varepsilon$.

Proof. $\quad T_\rho \hat{f}(S) = (1-\rho)^{|S|} \cdot \hat{f}(S)$