Sum of Squares Lower Bounds Versus Low-Degree Polynomial Lower Bounds

Aaron Potechin

University of Chicago

Outline

I. Introduction
II. Low-Degree Polynomial Lower Bound $\Leftrightarrow \tilde{E}[1]$ is well-behaved
III. Current Knowledge About Sum of Squares Lower Bounds
IV. Intuition for the Low-Degree Conjecture (time permitting)

Note: This talk is closely connected to Prasad Raghavendra's 4th bootcamp talk but is from a different perspective (looking at the current gaps between low-degree polynomial lower bounds and sum of squares lower bounds).

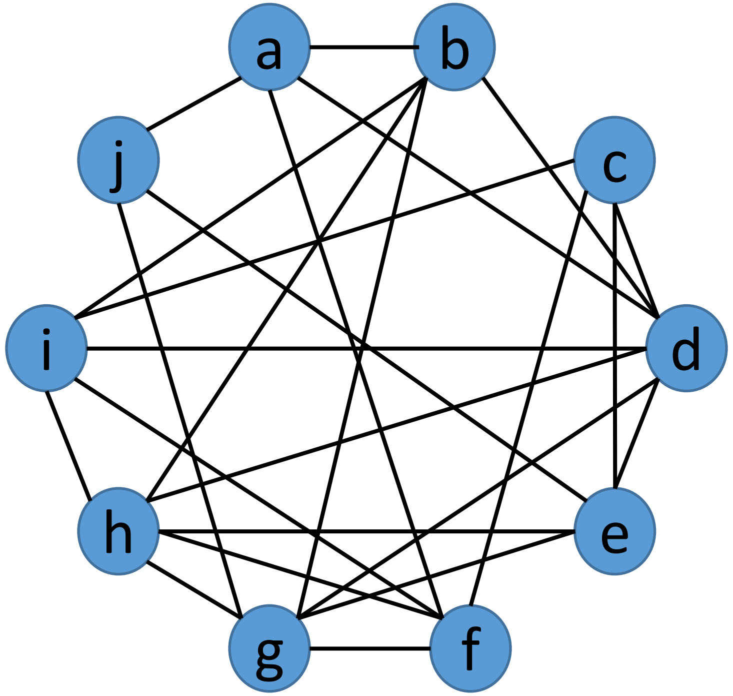
Part I: Introduction

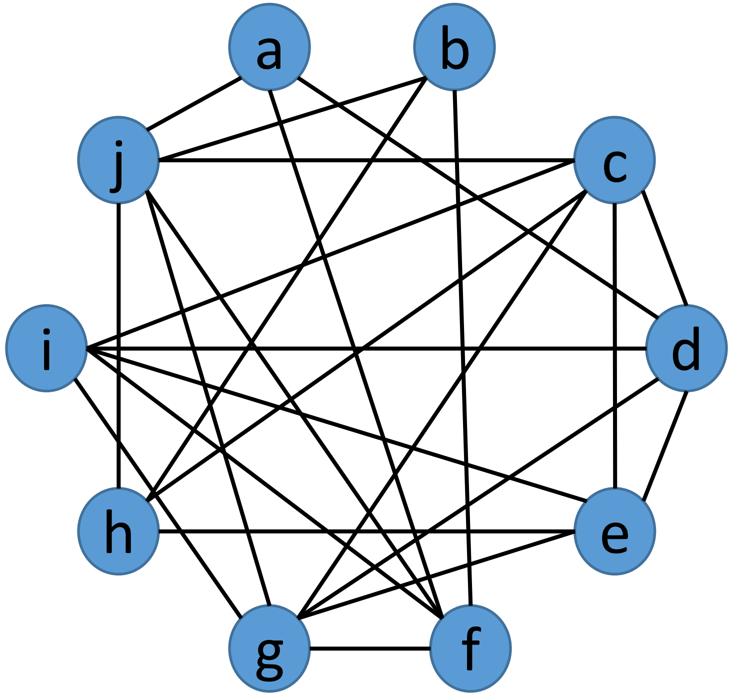
Distinguishing Problems

- Distinguishing problems: Given a random distribution D_{random} and a planted distribution $D_{planted}$, can we distinguish between these two distributions?
- Example: Planted clique:
  - D_{random}: $G\left(n, \frac{1}{2}\right)$
  - $D_{planted}$: $G\left(n, \frac{1}{2}\right)$ + clique of size k
- Example: Tensor PCA (principal component analysis):
  - D_{random}: $T_{i_1 \dots i_k} = N(0,1)$ (where k is the order of the tensor).
  - $D_{planted}$: $T_{i_1 \dots i_k} = N(0,1) + \lambda v_{i_1} v_{i_2} \dots v_{i_k}$ where $\lambda > 0$ and v is a unit vector.

Planted Clique Example

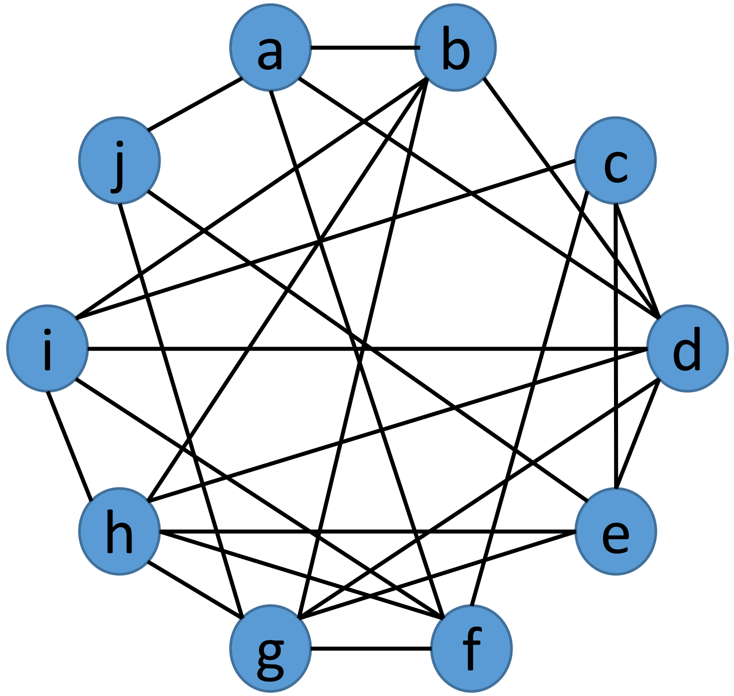
- Random instance: $G\left(n, \frac{1}{2}\right)$
- Planted instance: $G\left(n, \frac{1}{2}\right) + K_k$
- Example: Which graph has a planted 5-clique?

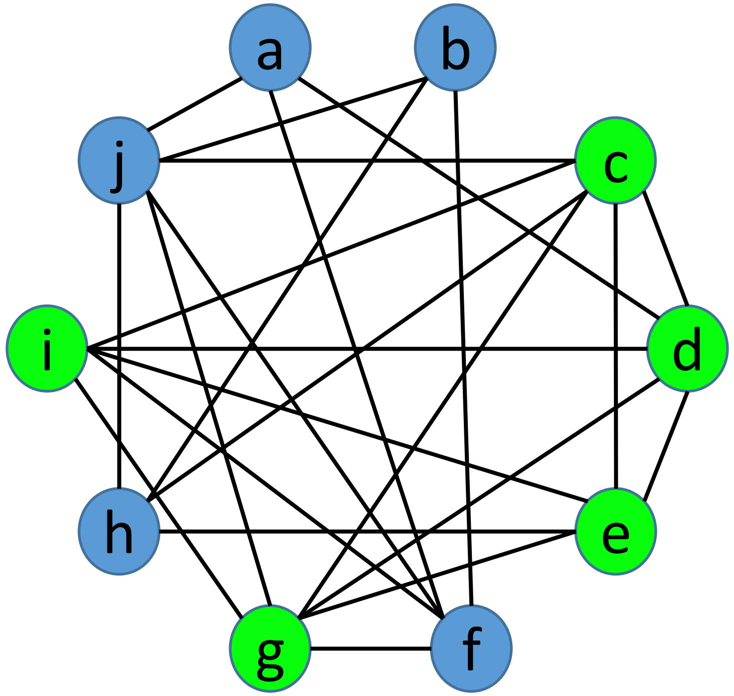




Planted Clique Example

- Random instance: $G\left(n, \frac{1}{2}\right)$
- Planted instance: $G\left(n, \frac{1}{2}\right) + K_k$
- Example: Which graph has a planted 5-clique?





Low-Degree Polynomial Framework

- Low-Degree Polynomial Framework: Is there a low-degree polynomial f which distinguishes between D_{random} and $D_{planted}$?
- More precisely, is there a low-degree polynomial f such that
  1. $E_{planted}[f]$ is large.
  2. $E_{random}[f] = 0$ and $E_{random}[f^2] \leq 1$.

  ?
- If there is no such polynomial f then we have a low-degree polynomial lower bound.

Sum of Squares (SoS) Framework

- The sum of squares hierarchy (SoS) is most naturally applied to certification problems (i.e. certifying that a random input does not have some hidden structure).
- That said, we can analyze distinguishing problems using the pseudo-calibration framework [BHK+16]:
  1. Use pseudo-calibration to obtain pseudo-expectation values for the random inputs.
  2. Construct the corresponding moment matrix M.
  3. Analyze whether $M \succcurlyeq 0$.
- If $M \succcurlyeq 0$ w.h.p. then we have an SoS lower bound.
- More precisely, the pseudo-expectation valules $\tilde{E}$ will satisfy all low-degree constraints satisfied by the planted distribution.

Summary

Start with a random and planted distribution.

Show that there is no low-degree polynomial f such that

1. $E_{planted}[f]$ is large
2. $E_{random}[f] = 0$ and $E_{random}\left[f^2\right] \leq 1$

Low-degree polynomial lower bound

Use pseudo-calibration to obtain pseudo-expectation values $\tilde{E}$.

Construct the corresponding moment matrix M.

Show $M \succcurlyeq 0$ w.h.p.

SoS lower bound

Low-Degree Conjecture

- SoS lower bound (where $\tilde{E}[1]$ is well-behaved) $\Rightarrow$ low-degree polynomial lower bound
- Low-degree conjecture: For symmetric distinguishing problems,

Low-degree polynomial lower bound $\Rightarrow$ SoS lower bound for a noisy version of the problem (where we add some additional noise to the planted distribution).

Part II: Low-Degree Polynomial Lower Bound $\Leftrightarrow$ $\tilde{E}[1]$ is well-behaved

Low-Degree Polynomial Lower Bound $\Leftrightarrow \tilde{E}[1]$ is well-behaved

- Observation on p. 71 in Sam Hopkin's thesis: $\tilde{E}[1]$ is the low-degree likelihood ratio for the input being from the planted distribution.
- What we'll show here: If there is a low-degree polynomial f such that
  1. $E_{planted}[f] = C$
  2. $E_{random}[f] = 0$ and $E_{random}[f^2] \leq 1$

  then $Var(\tilde{E}[1]) \geq C^2$.

Background: Fourier Analysis and Low-Degree Projections

- Setup: We have
  - A vector space of polynomials
  - An inner product $\langle f, g \rangle = E_{random}[fg]$
  - An orthonormal basis of Fourier characters $\{\chi_i\}$ which are polynomials.
- Fourier decomposition: For any polynomial f, we can write $f = \sum_{\chi_i} \hat{f}_i \chi_i$ where $\hat{f}_i = \langle f, \chi_i \rangle = E_{random}[f \chi_i]$.
- Low-degree projection: The low-degree projection of f is

$$\sum_{low\ degree\ \chi_i} \hat{f}_i \chi_i = \sum_{low\ degree\ \chi_i} E_{random}[f \chi_i] \chi_i$$

Goal: Assigning Pseudo-expectation Values

- Setup: We have
  - Solution variables for the planted structure.
  - Fourier characters χ_i on the random input
- Example: For the planted clique problem, we have
  - Solution variables x_i where we want that $x_i = 1$ if vertex i is in the planted clique and 0 otherwise.
  - Fourier characters $X_E = (-1)^{|E \setminus E(G)|} = \prod_{e \in E} \chi_{\{e\}}$ where $\chi_{\{e\}} = 1$ if $e \in E(G)$ and -1 otherwise.
- Each planted instance assigns values to the solution variables (and thus any polynomial p in the solution variables).
- Q: Given a random instance I, can we assign a pseudo-expectation value $\tilde{E}[p](I)$ to each low-degree polynomial p in the solution variables?

Pseudo-Calibration

- Pseudo-calibration: Take $\tilde{E}[p](I)$ to be the low-degree projection of

$$\frac{\Pr_{planted}(I)}{\Pr_{random}(I)} p(I)$$

- Reason: For any low-degree Fourier character χ_i,

$$\mathrm{E}_{random}\left[\tilde{E}[p](I)\chi_i\right] = E_{random}\left[\frac{\Pr_{planted}(I)}{\Pr_{random}(I)} p(I)\chi_i\right] = E_{planted}[p(I)\chi_i]$$

- Pseudo-calibration equation:

$$\tilde{E}[p](I) = \sum_{low-degree\ \chi_i} E_{planted}[p(I)\chi_i]\,\chi_i$$

Canonical Example: Planted Clique

- Random distribution: $G(n, 1/2)$
- Planted distribution: Start with a $G(n, 1/2)$ graph and put each vertex in the planted clique with probability k/n.
- Define $x_V = \prod_{i \in V} x_i$
- Claim: $E_{planted}[x_V \chi_E] = \left(\frac{k}{n}\right)^{|V \cup V(E)|}$ where $V(E)$ is the set of endpoints of edges in E.
- Reason:
  - If every vertex in $V \cup V(E)$ is in the planted clique then $x_V = 1$ and $\chi_E = 1$.
  - If some vertex in V is not in the planted clique then $x_V = 0$.
  - If some vertex in $V(E)$ is not in the planted clique then $E[\chi_E] = 0$ (where the expectation is over the part of G outside of the planted clique)
- Pseudo-expectation values: $\tilde{E}[x_V] = \sum_{E:|V \cup V(E)| \leq t} \left(\frac{k}{n}\right)^{|V \cup V(E)|} \chi_E$

Analyzing $\tilde{E}[1]$

- Pseudo-calibration equation: $\tilde{E}[p](I) = \sum_{low\ degree\ \chi_i} E_{planted}[p(I)\chi_i]\, \chi_i$
- Special case: $\tilde{E}[1] = 1 + \sum_{non-empty\ low\ degree\ \chi_i} E_{planted}[\chi_i]\, \chi_i$
- Assume we have a low-degree polynomial f such that
  - $E_{planted}[f] = C$
  - $E_{random}[f] = 0$ and $Var(f) \leq 1$
- Note: All sums below are over low-degree, non-empty χ_i.
- Write $f = \sum_{\chi_i} a_i \chi_i$ and let $b_i = E_{planted}[\chi_i]$. $\tilde{E}[1] - 1 = \sum_{\chi_i} b_i\, \chi_i$ so $Var(f) = \sum_{\chi_i} a_i^2$ and $\mathrm{Var}\left(\tilde{E}[1]\right) = \sum_{\chi_i} b_i^2$.
- Using Cauchy-Schwarz,

$$C = E_{planted}[f] = \sum_{\chi_i} a_i b_i \leq \sqrt{\sum_{\chi_i} a_i^2}\, \sqrt{\sum_{\chi_i} b_i^2} = \sqrt{Var(f) Var(\tilde{E}[1])}$$

- Thus, $Var\left(\tilde{E}[1]\right) \geq C^2$

Low-Degree Polynomial Lower Bounds Versus SoS Lower Bounds

Low-degree polynomial lower bound

SoS lower bound

$$\begin{array}{c|c|c|c|}
 & 1 & x_k & x_k x_l \\
\hline
1 & \tilde{E}[1] & \tilde{E}[x_k] & \tilde{E}[x_k x_l] \\
\hline
x_i & \tilde{E}[x_i] & \tilde{E}[x_i x_k] & \tilde{E}[x_i x_k x_l] \\
\hline
x_i x_j & \tilde{E}[x_i x_j] & \tilde{E}[x_i x_j x_k] & \tilde{E}[x_i x_j x_k x_l] \\
\hline
\end{array} \succcurlyeq 0$$

Moment matrix M

Summary

- SoS lower bounds using pseudo-calibration are strictly stronger than low-degree polynomial lower bounds as they involve analyzing the entire moment matrix.
- There are many interesting techniques involved in proving SoS lower bounds.
- That said, low-degree polynomials are an excellent heuristic for determining the computational threshold where a problem is hard and it is much easier to prove low-degree polynomial lower bounds.

Part III: Current Knowledge About Sum of Squares Lower Bounds

Evidence for the Low-Degree Conjecture

- The thresholds for SoS lower bounds and low-degree polynomials lower bounds match for
  - Planted clique [BHK+16]
  - Tensor PCA [HKP+17, PR20]
  - Sparse PCA [HKP+17, DNS20, PR20]
  - Random CSPs [KMOW17]
- However, there are still significant gaps between known SoS lower bounds and known low-degree polynomial lower bounds.

Delicateness of Current SoS Lower Bounds

- Subtle issue: Current SoS lower bound techniques are sensitive to the choice of planted distribution.
- Example: Planted Clique
  - Random distribution: $G\left(n, \frac{1}{2}\right)$
  - Planted distribution used in [BHK+16]: Put each vertex in the planted clique independently with probability $\frac{k}{n}$.
  - Desired planted distribution: Plant a clique of size exactly k.
  - For planted clique, Shuo Pang [P21] recently fixed this issue by proving an SoS lower bound for the desired planted distribution.

Delicateness of Current SoS Lower Bounds

- Subtle issue: Current SoS lower bound techniques are sensitive to the choice of planted distribution.
- Example: Tensor PCA
  - Random distribution: Tensor T with Gaussian entries
  - Planted distribution used in [HKP+17] and [PR20]: $T + \lambda(v \otimes v \otimes \cdots \otimes v)$ where v is a vector where each coordinate is in $\{-\frac{1}{\sqrt{\Delta n}}, 0, \frac{1}{\sqrt{\Delta n}}\}$ with probabilities $\frac{\Delta}{2}, 1-\Delta, \frac{\Delta}{2}$ where $\Delta = n^{-\epsilon}$.
  - If we instead take v to be a unit vector with coordinates $\pm\frac{1}{\sqrt{n}}$, the current techniques for analyzing the moment matrix M don't quite work.

Example: Parallel Pancakes

- Consider the following random and planted distribtions.
- Random: m random vectors $d_1, \ldots, d_m \in \mathbb{R}^n$ with $N(0,1)$ entries.
- Planted: First choose a unit vector $v \in \mathbb{R}^n$ with $\pm\frac{1}{\sqrt{n}}$ entries. Then choose m random vectors $d_1, \ldots, d_m \in \mathbb{R}^n$ with $N(0,1)$ entries and $a_1, \ldots, a_m$ from some distribution A and replace d_i with $d_i - \langle v, d_i \rangle v + a_i v$.
- In other words, $\langle d_i, v \rangle$ has distribution A and d_i is Gaussian in the directions orthogonal to the hidden direction v.
- Statistical query lower bound [DKS17]: If A matches the first k moments of $N(0,1)$ and $d_{TV}\big(A, N(0,1)\big) < \infty$ then there is a statistical query lower bound for $m \ll n^{\frac{k+1}{2}}$.

Special Case: $A = \{-1,1\}$

- For the special case when $A = \{-1,1\}$, we have an SoS lower bound for $m \ll n^{3/2}$ which was used to prove an SoS lower bound for the Sherrington-Kirkpatrick problem [GJJ+20].
- Note: There is a low-degree polynomial lower bound when $m \ll n^2$.
- Open problem: Can we strengthen the SoS lower bound from $m \ll n^{3/2}$ to $m \ll n^2$?
- Open problem: Can we prove SoS lower bounds for more general distributions A?

Example: Independent Set on Sparse Graphs

- Q: Given a sparse graph G with average degree $\approx d$, does it have an independent set of size $\approx k = \frac{n}{d^{1/2+\epsilon}}$?
- Random Distribution: Random $G\left(n, \frac{d}{n}\right)$ graph
- Naïve Planted Distribution: Start with a random $G\left(n, \frac{d}{n}\right)$ graph and put each vertex in the independent set with probability $\frac{k}{n}$.
- Problem: It is easy to distinguish these distributions! In fact, counting the number of edges is sufficient. This can be fixed by starting with a $G\left(n, \frac{d'}{n}\right)$ graph instead of a $G\left(n, \frac{d}{n}\right)$ graph for $d' = d\left(\frac{n^2}{n^2-k^2}\right)$, but then counting the number of triangles is still sufficient.
- What can we do?

Example: Independent Set on Sparse Graphs

- Low-degree polynomial lower bound for recovery [SW20]: Even though it is easy to distinguish the random and planted distributions, there is no low-degree polynomial which approximates the indicator function for whether a given vertex i is in the independent set.
- SoS certification lower bound [JPR+21]: We can tweak the pseudo-expectation values given by pseudo-calibration to show an SoS lower bound on the certification problem of proving that a $G\left(n, \frac{d}{n}\right)$ graph does not have an independent set of size $\approx k$.
- Note: To do this, we ignore all shapes α which have a component which is disconnected from $U_\alpha \cup V_\alpha$, which corresponds to ignoring all of the global distinguishers.

Open Problem: Quiet Planting

- Q: Can we find a planted distribution for independent set on sparse graphs which is hard to distinguish from $G\left(n, \frac{d}{n}\right)$ (or alternatively, from a random d-regular graph on n vertices)?

Part IV: Intuition for the Low-Degree Conjecture

Example: Maximum Eigenvalue of a Random Matrix

- Q: Given a symmetric matrix M, is $\lambda_{max}(M) \geq 2\sqrt{n} + 2$?
- Random distribution: A random symmetric $n \times n$ matrix M with Gaussian entries
- Planted distribution:
  1. Start with a random matrix M.
  2. Letting v be the eigenvector of M with the largest eigenvalue, take $M' = M + \left(2\sqrt{n} + 2 - \lambda_{max}(M)\right) vv^T$.
- Note: For a random symmetric $n \times n$ matrix M with Gaussian entries, w.h.p. $\lambda_{max}(M)$ is $2\sqrt{n} + O\left(\frac{1}{n^{1/6}}\right)$ and is described by the Tracy-Widom distribution [TW94].

Example: Maximum Eigenvalue of a Random Matrix

- Q: Given a symmetric matrix M, is $\lambda_{max}(M) \geq 2\sqrt{n} + 2$?
- By its nature, SoS easily solves this problem.
- For any symmetric matrix M, $\lambda_{max}(M)Id - M \succcurlyeq 0$ so $\mathrm{x}^{\mathrm{T}}(\lambda_{max}(M)Id - M)x$ is a sum of squares which certifies that for any vector x, $x^T M x \leq \lambda_{max}(M)\|x\|^2$.
- However, since the planted distribution is only a slight tweak of the random distribution, this is very hard for low-degree polynomials to detect.
- Note: This example is delicate. For example, if we instead ask whether $\lambda_{max}(M) \geq C\sqrt{n}$ then low-degree polynomials can solve this problem via the trace power method.

Spectral Distinguishers

- Recall: A low-degree polynomial distinguisher is a polynomial f such that
  1. $E_{planted}[f]$ is large.
  2. $E_{random}[f] = 0$ and $E_{random}[f^2] \leq 1$.
- A spectral distinguisher is a matrix Q such that such that
  1. Each entry of Q is a low-degree polynomial in the entries of the input.
  2. $E_{planted}[\lambda^+_{max}(Q)]$ is large.
  3. $E_{random}[\lambda^+_{max}(Q)] \leq 1$.

  where $\lambda^+_{max}(Q)$ is the largest positive eigenvalue of Q and is 0 if $Q \preccurlyeq 0$.
- [HKP+17]: If SoS succeeds at a noisy version of the distinguishing problem (and certain technical conditions are satisfied) then there is a spectral distinguisher.

Spectral Distinguisher Example

- For the maximum eigenvalue problem, we can take
$$Q = C\left(M - \left(2\sqrt{n} + 1\right)Id\right)$$
- In the planted case, $\lambda_{max}(M) \geq 2\sqrt{n} + 2$ so $\lambda_{max}^{+}(Q) \geq C$.
- In the random case, w.h.p. $\lambda_{max}(M) = 2\sqrt{n} + O\left(\frac{1}{n^{1/6}}\right)$ so $\lambda_{max}^{+}(Q) = 0$. Thus, $E_{random}[\lambda_{max}^{+}(Q)]$ is very small.

Path for Proving the Low-Degree Conjecture

- Likely strengthening of this result: If SoS solves a noisy version of the distinguishing problem then there is a matrix M such that
  1. Each entry of M is a low-degree polynomial in the entries of the input.
  2. $E_{planted}[||M||]$ is large.
  3. $P_{random}(||M|| > 1)$ is very small.
- If so, then $tr\left(\left(MM^T\right)^q\right)$ is a low-degree distinguisher for $q = O(logn)$.

Thank You!