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

Aaron Potechin

University of Chicago

Background for this Talk

- The framework for proving SoS lower bounds on average case problems was pioneered by “A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem” by Boaz Barak, Sam Hopkins, Jonathan Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin [BHKKMP16].
- This paper was a major inspiration for the low-degree polynomial framework for analyzing average case problems.
- Sam Hopkin’s PhD thesis [Hop18] is a very good reference for the material in this talk.

Outline

I. Overview
II. Analyzing Low-Degree Polynomials
III. The Sum of Squares Hierarchy
IV. Pseudo-calibration
V. Low-Degree Polynomial Lower Bound $\Leftrightarrow$ $\tilde{E}[1]$ is Well-behaved
VI. Graph Matrices
VII. Current Sum of Squares Lower Bounds for Average Case Problems

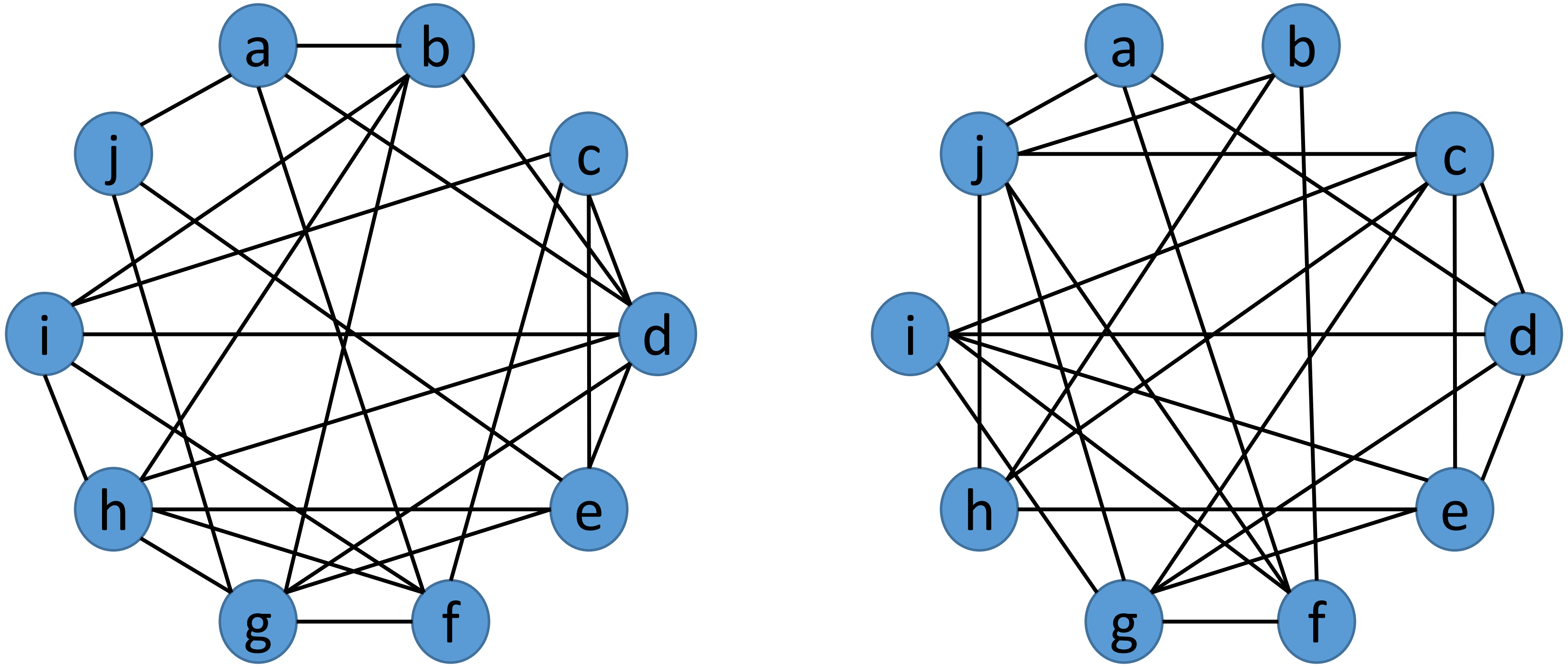
Part I: Overview

Distinguishing/Hypothesis Testing Problems

- Distinguishing problems: Given a random distribution and a planted distribution, can we distinguish between these two distributions?
- Example: Planted Clique
  - Random distribution: $G\left(n, \frac{1}{2}\right)$
  - Planted distribution: $G\left(n, \frac{1}{2}\right) +$ clique of size k
- Example: Non-Gaussian Component Analysis (NGCA)
  - Random distribution: m samples from $N(0, Id_n)$.
  - Planted distribution: First choose a random unit direction $\vec{v} \in R^n$. Then take m samples which have some distribution A in direction $\vec{v}$ and have distribution $N(0,1)$ in directions orthogonal to $\vec{v}$.

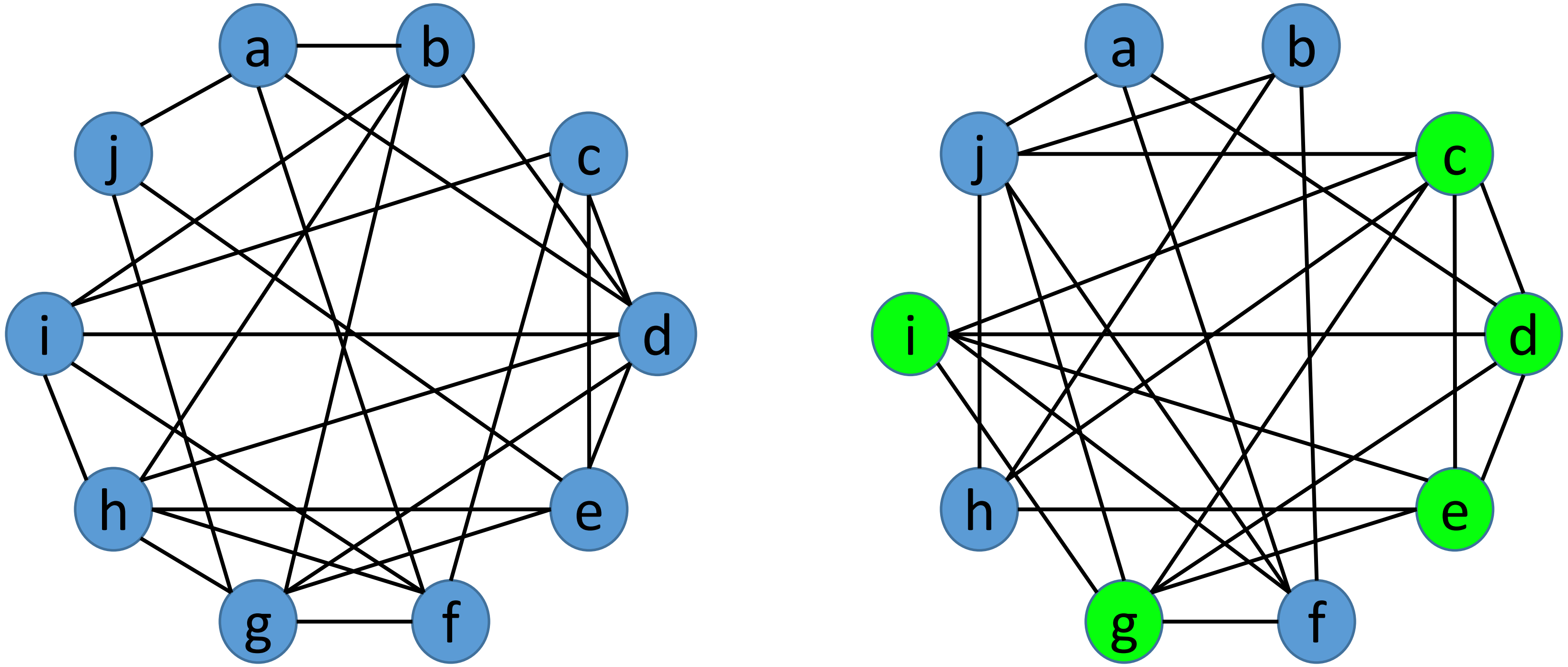
Planted Clique Example

- 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 $E_{planted}[f]$ is large, $E_{random}[f] = 0$, 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 [BHKKMP16]:
  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 $E_{random}[f^2] \leq 1$

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

- Fact: SoS lower bound proved via pseudo-calibration (where $\tilde{E}[1]$ is well-behaved) ⇒ low-degree polynomial lower bound
- Low-degree conjecture (see [Hop18] and [HW21]): For symmetric distinguishing problems, if there is a low-degree polynomial lower bound then no polynomial time algorithm can solve a noisy version of the problem where we add some additional noise to the planted distribution.
- SoS version of the low-degree conjecture: For symmetric distinguishing problems, if there is a low-degree polynomial lower bound then there is an SoS lower bound for a noisy version of the problem where we add some additional noise to the planted distribution.

Part II: Analyzing Low-Degree Polynomials

Analyzing Low-Degree Polynomials

- Key question: Is there a low-degree polynomial f such that $E_{planted}[f]$ is large, $E_{random}[f] = 0$, and $E_{random}[f^2] \leq 1$?
- This can be analyzed using the low-degree likelihood ratio (see e.g. [Hop18], [KWB22]). We will instead give a direct analysis.

Fourier Analysis on Random Inputs

- Setup: Assume that we have
  - A vector space of polynomials of the input entries.
  - An inner product $\langle f, g \rangle = E_{random}[fg]$.
  - An orthonormal basis of Fourier characters $\{\chi_E\}$ where $\chi_\emptyset = 1$.
- Example: $G(n, 1/2)$
  - We have the inner product $\langle f, g \rangle = E_{G \sim G(n,1/2)}[f(G)g(G)]$.
  - We have the Fourier characters $\chi_E(G) = (-1)^{|E \setminus E(G)|} = \prod_{e \in E} \chi_{\{e\}}(G)$ where $\chi_{\{e\}(G)} = 1$ if $e \in E(G)$ and -1 if $e \notin E(G)$.
  - This is essentially Fourier analysis over the Boolean hypercube where we have a variable for each potential edge.

Choosing the Best Low-Degree Polynomial

- Let $b_E = E_{planted}[\chi_E]$. Given a polynomial $f = \sum_E c_E \chi_E$, we have that
  - $E_{planted}[f] = \sum_E b_E c_E$
  - $E_{random}[f] = c_\emptyset$
  - $E_{random}\left[f^2\right] = \sum_E c_E^2$
- Goal: Find the polynomial f of degree at most d which maximizes $E_{planted}[f]$ subject to $E_{random}[f] = 0$ and $E_{random}\left[f^2\right] \leq 1$.
- Goal restatement: Maximize $\sum_{E:|E|\leq d} b_E c_E$ subject to $c_\emptyset = 0$ and $\sum_{E:0<|E|\leq d} c_E^2 \leq 1$.

Choosing the Best Low-Degree Polynomial Continued

- Let $b_E = E_{planted}[\chi_E]$. We want to maximize $\sum_{E:0<|E|\leq d} b_E c_E$ subject to $\sum_{E:0<|E|\leq d} c_E^2 \leq 1$.
- Claim: The maximum value of $\sum_{E:0<|E|\leq d} b_E c_E$ is $\sqrt{\sum_{E:0<|E|\leq d} b_E^2}$ which is achieved by taking $c_E = \frac{b_E}{\sqrt{\sum_{E:0<|E|\leq d} b_E^2}}$.
- Proof: By Cauchy Schwarz,

$$\sum_{E:0<|E|\leq d} b_E c_E \leq \sqrt{\sum_{E:0<|E|\leq d} b_E^2}\sqrt{\sum_{E:0<|E|\leq d} c_E^2} \leq \sqrt{\sum_{E:0<|E|\leq d} b_E^2}$$

- Taking $c_E = \frac{b_E}{\sqrt{\sum_{E:0<|E|\leq d} b_E^2}}$ gives $\sum_{E:0<|E|\leq d} b_E c_E = \sqrt{\sum_{E:0<|E|\leq d} b_E^2}$.

Analyzing Low-Degree Polynomials Summary

- The polynomial f of degree at most d which maximizes $E_{planted}[f]$ subject to $E_{random}[f] = 0$ and $E_{random}[f^2] \leq 1$ is

$$f = \frac{\sum_{E:0<|E|\leq d} E_{planted}[\chi_E]\chi_E}{\sqrt{\sum_{E:0<|E|\leq d}\left(E_{planted}[\chi_E]\right)^2}}$$

which gives $E_{planted}[f] = \sqrt{\sum_{E:0<|E|\leq d}\left(E_{planted}[\chi_E]\right)^2}$.

- If $\sum_{E:0<|E|\leq d}\left(E_{planted}[\chi_E]\right)^2 \gg 1$ then degree d polynomials can distinguish the random and planted distributions. If $\sum_{E:0<|E|\leq d}\left(E_{planted}[\chi_E]\right)^2$ is $o(1)$ then degree d polynomials do not distinguish the random and planted distributions.

Example: Planted Clique

- For planted clique, we can take the following random and planted distributions[1]:
  - Random distribution: $G(n, 1/2)$
  - Planted distribution: $G(n, 1/2)$ plus a planted clique where we put each vertex in the planted clique independently with probability k/n.
- We want to compute $\sum_{E:0<|E|\leq d}\left(E_{planted}[\chi_E]\right)^2$
- Claim: $E_{planted}[\chi_E] = \left(\frac{k}{n}\right)^{|V(E)|}$ where $V(E)$ is the set of endpoints of edges in E.
- Idea: For the planted distribution, if all of the vertices in $V(E)$ are in the planted clique then $\chi_E = 1$. Otherwise, $E[\chi_E] = 0$.

[1]Ideally, we'd like to use the planted distribution where the clique has size exactly k. We use this planted distribution to make the SoS lower bound analysis easier.

Low-Degree Analysis for Planted Clique

- We have that $E_{planted}[\chi_E] = \left(\frac{k}{n}\right)^{|V(E)|}$ and we want to compute $\sum_{E:0<|E|\leq d}\left(E_{planted}[\chi_E]\right)^2$.
- For each $j \in [2d]$, there are at most $2^{j^2/2}n^j$ different sets E such that $|V(E)| = j$.
- $\sum_{E:0<|E|\leq d}\left(E_{planted}[\chi_E]\right)^2 \leq \sum_{j=1}^{2d} 2^{\frac{j^2}{2}}\left(\frac{k^2}{n}\right)^j \leq \sum_{j=1}^{2d}\left(\frac{2^d k^2}{n}\right)^j$
- This is $o(1)$ as long as k is $o(n^{\frac{1}{2}-\frac{d}{2\log(n)}})$

Part III: The Sum of Squares Hierarchy

Setup for the Sum of Squares Hierarchy

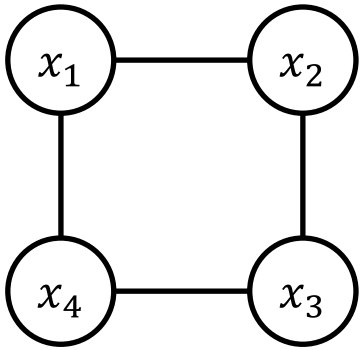
- This talk: We view the sum of squares hierarchy (SoS) as a proof system for determining whether or not a system of polynomial equations is feasible over the real numbers.
- Example: k-clique equations
  - For all $i \in [n]$, $x_i^2 = x_i$.
  - $x_i x_j$ = 0 if $\{i,j\} \notin E(G)$.
  - $\sum_{i=1}^{n} x_i = k$.
- These equations are feasible precisely when G contains a k-clique. If SoS can prove that these equations are infeasible then this certifies that G does not have a k-clique.

Positivstellensatz/Sum of Squares Proofs

- Given a system of polynomial equations $\{s_i = 0\}$ over R, a degree d Positivstellenstz/sum of squares proof of infeasibility is an equality of the form $-1 = \sum_i f_i s_i + \sum_j g_j^2$ where
  - For all i, $\deg(f_i) + \deg(s_i) \leq d$.
  - For all j, $\deg(g_j) \leq d/2$.

Positivstellensatz/Sum of Squares Proof Example

- Consider the following system of polynomial equations corresponding to the statement that C_4 has a triangle:
  1. For all $i \in [4]$, $x_i^2 - x_i = 0$.
  2. $x_1 x_3 = 0$ and $x_2 x_4 = 0$.
  3. $x_1 + x_2 + x_3 + x_4 - 3 = 0$.



- A degree 2 Positivstellensatz/SoS proof of infeasibility is as follows:

$$-1 = (x_1 + x_3 - 1)^2 + (x_2 + x_4 - 1)^2 - 2x_1 x_3 - 2x_2 x_4 - \sum_{i=1}^{4} \left(x_i^2 - x_i\right) + (x_1 + x_2 + x_3 + x_4 - 3)$$

Pseudo-expectation Values

- Given polynomial equalities $\{s_i = 0\}$, degree d pseudo-expectation values are a linear map $\tilde{E}$ from polynomials of degree at most d to R such that:
  - $\tilde{E}[1] = 1$.
  - $\tilde{E}[f s_i] = 0$ whenever $\deg(f) + \deg(s_i) \leq d$.
  - $\tilde{E}\left[g^2\right] \geq 0$ whenever $\deg(g) \leq d/2$.
- Proposition: We cannot have both degree d pseudo-expectation values $\tilde{E}$ and a degree d SoS/Positivstellensatz proof of infeasibility.
- Proof: Assume we have both. Applying the degree d pseudo-expectation values to the degree d SoS/Positivstellensatz proof of infeasibility $-1 = \sum_i f_i s_i + \sum_j g_j^2$ gives

$$-1 = \tilde{E}[-1] = \sum_i \tilde{E}[f_i s_i] + \sum_j \tilde{E}[g_j^2] \geq 0$$

which gives a contradiction.

Example: Knapsack with Unit Weights and Capacity k

- Equations: We have a variable x_i for each weight. We want that $x_i = 1$ if we take weight i and $x_i = 0$ otherwise. We can capture this with the following equations:
  - For all $i \in [n]$, $x_i^2 = x_i$.
  - $\sum_{i=1}^{n} x_i = k$.
- These equations are infeasible whenever $k \notin \mathbb{Z} \cap [0, n]$. SoS is poor at capturing integrality arguments so SoS requires degree $2\lceil \min\{k, n-k\} \rceil$ to refute these equations [Gri01a].
- Degree 2 pseudo-expectation values for $n = 3, k = 3/2$: $\tilde{E}\left[x_i^2\right] = \tilde{E}[x_i] = 1/2$ for all i, $\tilde{E}\left[x_i x_j\right] = 1/8$ whenever $i \neq j$.

Checking the Pseudo-expectation Values

- Equations:
  - For all $i \in [3]$, $x_i^2 = x_i$.
  - $\sum_{i=1}^{3} x_i = 3/2$.
- Pseudo-expectation values: $\tilde{E}[x_i^2] = \tilde{E}[x_i] = 1/2$ for all i, $\tilde{E}[x_i x_j] = 1/8$ whenever $i \neq j$.
- We can check that the polynomial equalities are satisfied as follows:
  - $\tilde{E}[x_1 + x_2 + x_3] = 1/2 + 1/2 + 1/2 = 3/2$.
  - $\tilde{E}[x_1^2 + x_1 x_2 + x_1 x_3] = 1/2 + 1/8 + 1/8 = 3/4 = (3/2)\tilde{E}[x_1]$.

The Moment Matrix

- To check that $\tilde{E}\left[g^2\right] \geq 0$ whenever $\deg(g) \leq d/2$, we can use the moment matrix M whose rows and columns are indexed by monomials of degree at most $d/2$ with entries $M_{pq} = \tilde{E}[pq]$.
- Fact: $\tilde{E}\left[g^2\right] \geq 0$ whenever $\deg(g) \leq d/2 \Leftrightarrow M \succcurlyeq 0$ (i.e., M is positive semidefinite).

Checking $M \succcurlyeq 0$

- Pseudo-expectation values: $\tilde{E}\left[x_i^2\right] = \tilde{E}[x_i] = 1/2$ for all i, $\tilde{E}\left[x_i x_j\right] = 1/8$ whenever $i \neq j$.
- The corresponding moment matrix is $M = \begin{pmatrix} 1 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/8 & 1/8 \\ 1/2 & 1/8 & 1/2 & 1/8 \\ 1/2 & 1/8 & 1/8 & 1/2 \end{pmatrix}$.
- To see that $M \succcurlyeq 0$, observe that

$$\begin{pmatrix} 1 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/8 & 1/8 \\ 1/2 & 1/8 & 1/2 & 1/8 \\ 1/2 & 1/8 & 1/8 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1/4 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 & 1/4 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/4 & -1/8 & -1/8 \\ 0 & -1/8 & 1/4 & -1/8 \\ 0 & -1/8 & -1/8 & 1/4 \end{pmatrix}.$$

SoS Lower Bounds

- Summary: To prove a degree d SoS lower bound, we generally need to
  1. Construct candidate degree d pseudo-expectation values $\tilde{E}$.
  2. Show that $\tilde{E}$ gives valid degree d pseudo-expectation values. The most difficult condition to check is that the moment matrix M is PSD (positive semidefinite).

Part IV: Pseudo-calibration

Proving SoS Lower Bounds for Average-Case Problems

- How can we prove SoS lower bounds for average case problems?
- Key idea from [BHKKMP16]: To show that degree d SoS fails to certify that no solution exists, show that degree d SoS fails to distinguish between
  1. The random input distribution (where there is no solution w.h.p.).
  2. A planted distribution which always has a solution.
- We can construct the pseudo-expectation values $\tilde{E}$ for the random input by using the planted distribution as a guide and using pseudo-calibration [BHKKMP16].

Pseudo-calibration

- Pseudo-calibration technique [BHKKMP16]: Construct $\tilde{E}$ so that for all low-degree tests, the behavior of $\tilde{E}$ on random inputs matches the behavior of actual solutions for the planted distribution.
- Pseudo-calibration equation: For all polynomials p of degree at most d and all small E (for an appropriate definition of small),
$$E_{random}\left[\tilde{E}[p]\chi_E\right] = E_{planted}[p\chi_E]$$
- This implies that for all such p and E, the Fourier coefficient $\widehat{\tilde{E}[p]}_E$ is $\widehat{\tilde{E}[p]}_E = E_{planted}[p\chi_E]$. If we take the other Fourier coefficients to be 0, we have that $\tilde{E}[p] = \sum_{small\ E} E_{planted}[p\chi_E]\chi_E$.

Pseudo-calibration Example: Planted Clique

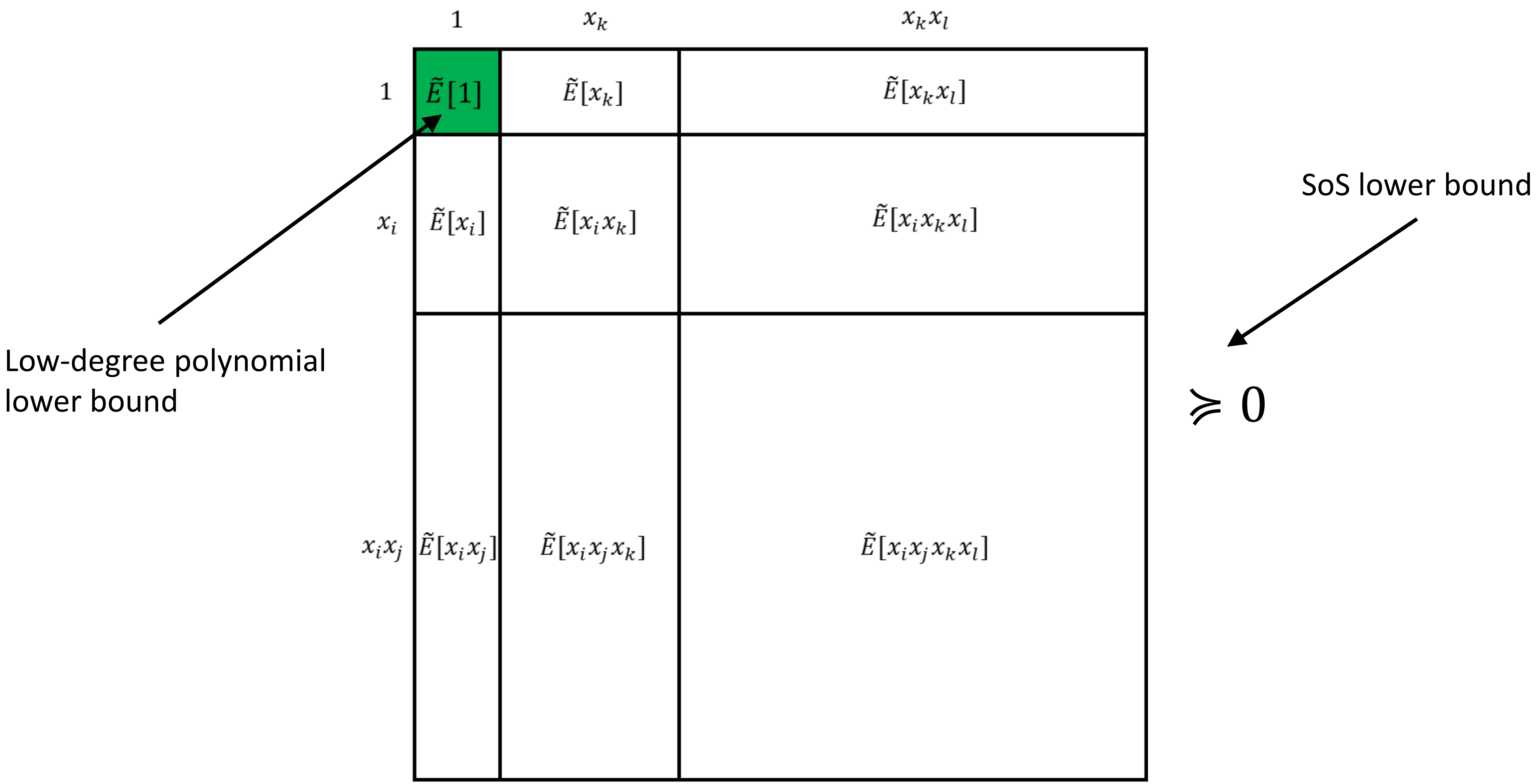
- Pseudo-calibration equation: $\tilde{E}[p] = \sum_{small\ E} E_{planted}[p\chi_E]\chi_E$
- Planted clique distributions:
  - Random distribution: $G(n, 1/2)$.
  - Planted distribution: $G(n, 1/2)$ plus a planted clique where we put each vertex in the planted clique independently with probability k/n.
- Definition: 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.
- Pseudo-expectation values: $\tilde{E}[x_V] = \sum_{E:|V \cup V(E)| \leq t} \left(\frac{k}{n}\right)^{|V \cup V(E)|} \chi_E$

Part V: Low-Degree Polynomial Lower Bound $\Leftrightarrow Var\left(\tilde{E}[1]\right)$ is $o(1)$

Analyzing $Var\left(\tilde{E}[1]\right)$

- Using pseudo-calibration gives $\tilde{E}[p] = \sum_{small\ E} E_{planted}[p\chi_E]\chi_E$.
- Special case: $\tilde{E}[1] = 1 + \sum_{E:E\ is\ small,\ E\neq\emptyset} E_{planted}[\chi_E]\chi_E$.
- $Var\left(\tilde{E}[1]\right) = E_{random}\left[\left(\sum_{E:E\ is\ small,\ E\neq\emptyset} E_{planted}[\chi_E]\chi_E\ \right)^2\right] =$ $E_{random}\left[\sum_{E,E':E,E'\ are\ small,\ E\neq\emptyset,E'\neq\emptyset} E_{planted}[\chi_E]E_{planted}\left[\chi_{E'}\right]\chi_E\chi_{E'}\ \right]$ $= \sum_{E:E\ is\ small,\ E\neq\emptyset}\left(E_{planted}[\chi_E]\right)^2$.
- This is the same expression we analyzed for low-degree polynomials!
- Corollary: Low-Degree Polynomial Lower Bound $\Leftrightarrow$ $Var\left(\tilde{E}[1]\right)$ is $o(1)$

Low-Degree Polynomial Lower Bounds Versus SoS Lower Bounds



Moment matrix M

Summary

- SoS lower bounds proved via 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 for where a problem is hard and it is much easier to prove low-degree polynomial lower bounds.

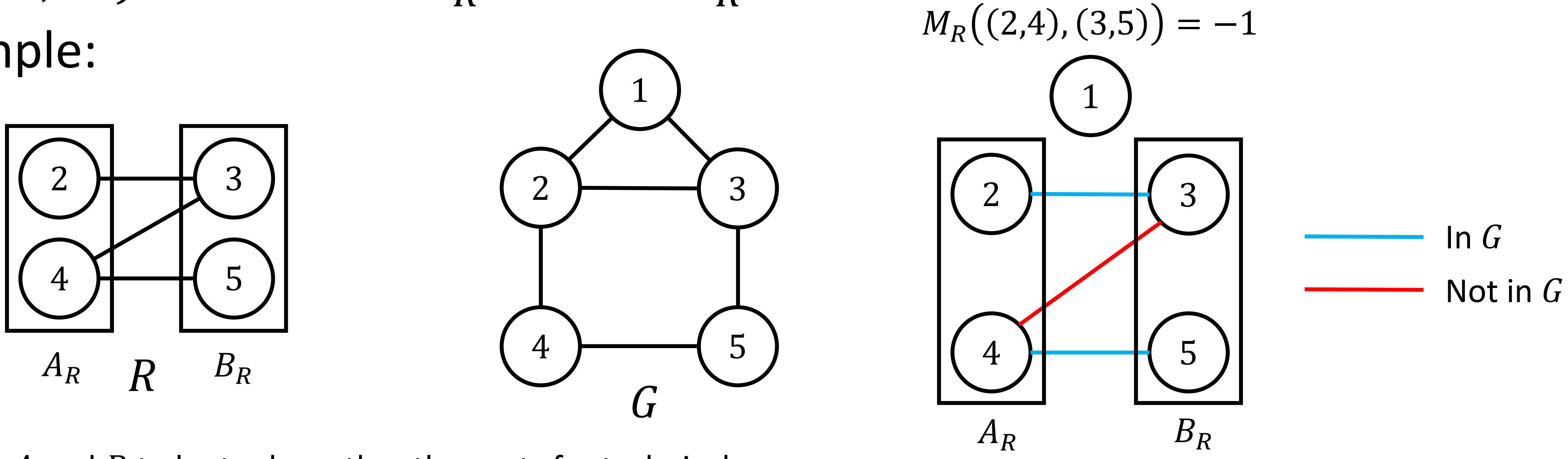
Part VI: Graph Matrices

Background on Graph Matrices

- Graph matrices are a type of matrix which is a key technical tool for analyzing SoS on average case problems.
- Recently, graph matrices have been used to analyze power-sum decompositions of polynomials [BHKX22], to analyze the ellipsoid fitting conjecture [PTVW23, HKPX23], and to analyze a class of first-order iterative algorithms including belief propagation and approximate message passing [JP24].
- Currently, not that much is known about graph matrices except for rough norm bounds [AMP20, JPRTX21, RT23].
- The limiting distribution of the spectrum of the singular values as $n \to \infty$ (i.e., an analogue of Wigner's Semicircle Law) was determined for one family of graph matrices called multi-Z-shaped graph matrices [CP20, CP22].

Ribbons

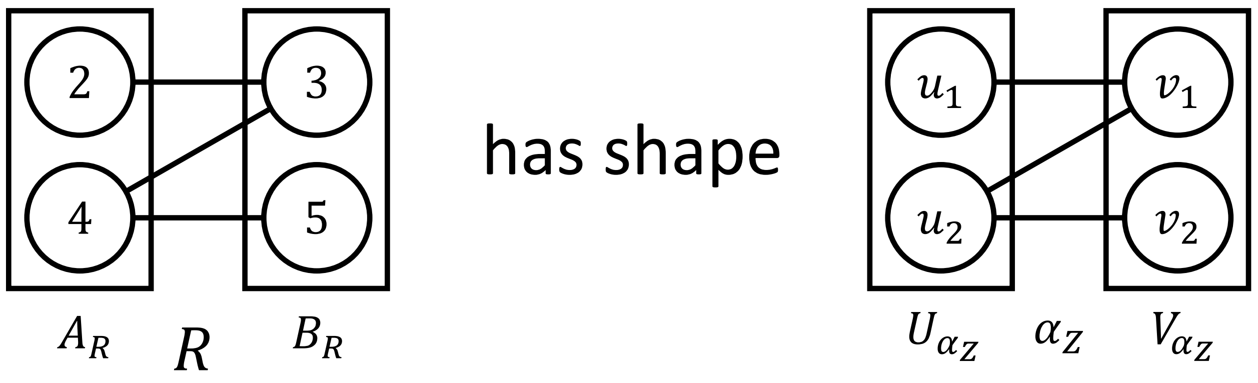
- Definition: We define a ribbon to consist of a set of edges $E(R)$ together with distinguished tuples[1] A_R and B_R of elements in $[n]$. We call A_R and B_R the left and right sides of R.
- We take M_R to be the matrix where $M_R(A_R, B_R) = \chi_{E(R)}(G)$ and $M_R(A', B') = 0$ if $A' \neq A_R$ or $B' \neq B_R$.
- Example:



[1]We take A and B to be tuples rather than sets for technical reasons.

Shapes

- Definition: A shape α consists of a graph α with distinguished tuples of vertices U_α and V_α which we call the left and right sides of α.
- Definition: We say that a ribbon R has shape α if there is an injective map $\sigma: V(\alpha) \to [n]$ such that $\sigma(\alpha) = R$. More precisely, $\sigma(U_\alpha) = A_R$, $\sigma(V_\alpha) = B_R$, and $\sigma\big(E(\alpha)\big) = E(R)$.
- Example:



Graph Matrices

- Recall: Given a ribbon R, M_R is the matrix where $M_R(A,B) = \chi_{E(R)}(G)$ and $M_R(A',B') = 0$ if $A' \neq A$ or $B' \neq B$.
- Definition: Given a shape α, the graph matrix M_α is

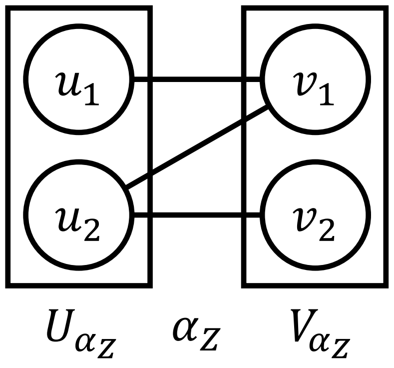
$$M_\alpha = \sum_{Ribbons\ R\ of\ shape\ \alpha} M_R$$

- Equivalently, $M_\alpha(A,B) = \frac{1}{|Aut(\alpha)|} \sum_{\substack{\sigma:V(\alpha)\to V(G):\\ \sigma\ is\ injective,\\ \sigma(U_\alpha)=A,\sigma(V_\alpha)=B}} \chi_{\sigma(E(\alpha))}(G)$

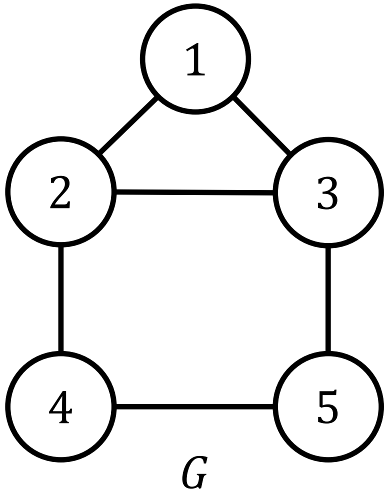
  where $Aut(\alpha)$ is the set of automorphisms of α which keep U_α and V_α fixed.
- Note that M_α is a $\frac{n!}{(n-|U_\alpha|)!} \times \frac{n!}{(n-|V_\alpha|)!}$ matrix with rows and columns indexed by tuples A and B of $|U_\alpha|$ and $|V_\alpha|$ elements respectively.

Example: Z-Shaped Graph Matrix

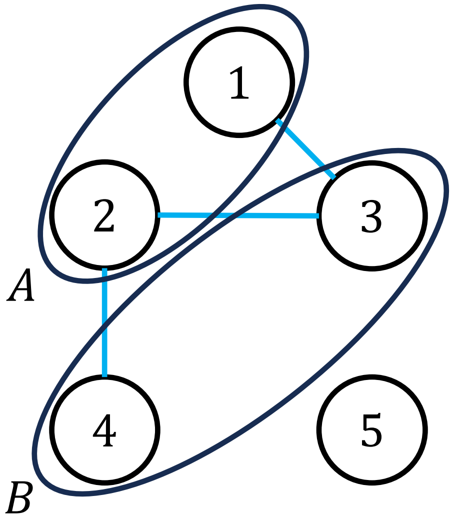
- $M_{\alpha_Z} = \sum_{Ribbons\ R\ with\ shape\ \alpha_Z} M_R.$
- $M_{\alpha_Z}(A,B) = \sum_{\substack{\sigma:V(\alpha_Z)\to V(G): \\ \sigma\ is\ injective, \\ \sigma(U_{\alpha_Z})=A, \sigma(V_{\alpha_Z})=B}} \chi_{\sigma(E(\alpha_Z))}(G).$



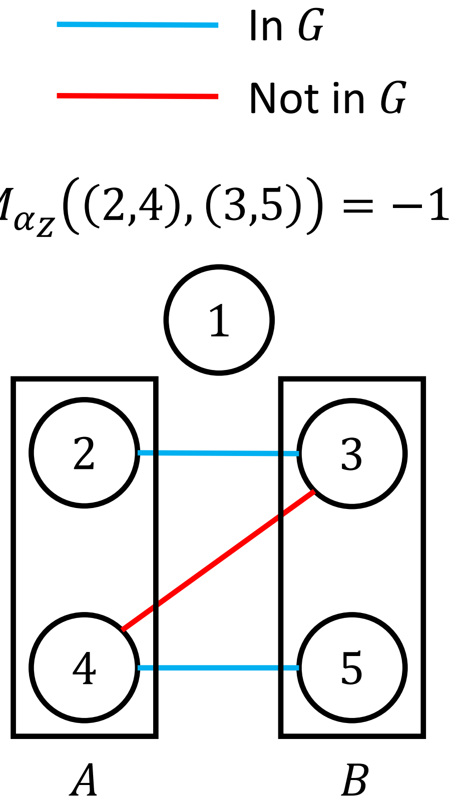
Some entries of M_{α_Z} for a given input graph G:



$M_{\alpha_Z}((1,2),(3,4)) = 1$



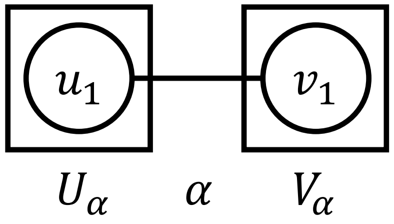
$M_{\alpha_Z}((2,4),(3,5)) = -1$

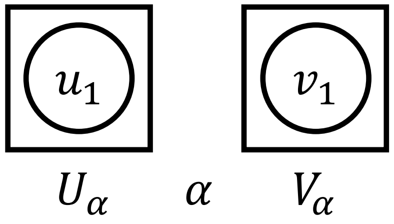


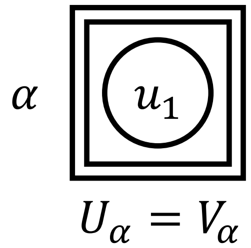
More Graph Matrix Examples

- Graph matrix examples (for these examples, $V(\alpha) = U_\alpha \cup V_\alpha$):
  1. If α is the shape with $U_\alpha = (u_1)$, $V_\alpha = (v_1)$, and $E(\alpha) = \{\{u_1, v_1\}\}$ then M_α is a symmetric random matrix with ± 1 entries and 0s on the diagonal.
  2. If α is the shape with $U_\alpha = (u_1)$, $V_\alpha = (v_1)$, and $E(\alpha) = \{\}$ then $M_\alpha = J - Id$ where J is the all ones matrix.
  3. If α is the shape with $U_\alpha = V_\alpha = (u_1)$, and $E(\alpha) = \{\}$ then $M_\alpha = Id$

$$\begin{pmatrix} 0 & \pm 1 & \pm 1 & \pm 1 \\ \pm 1 & 0 & \pm 1 & \pm 1 \\ \pm 1 & \pm 1 & 0 & \pm 1 \\ \pm 1 & \pm 1 & \pm 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$



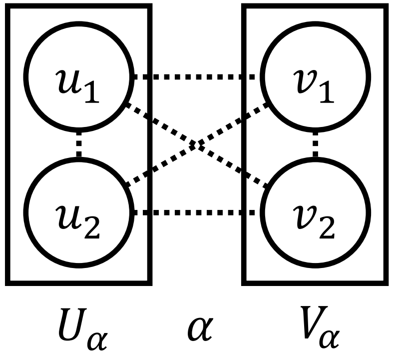




Example: Decomposing a Clique Indicator Matrix

- Let M be the $n(n-1) \times n(n-1)$ clique indicator matrix with entries $M\big((a,b),(c,d)\big) = 1$ if $\{a,b,c,d\}$ is a 4-clique and 0 otherwise.
- Using graph matrices, we can decompose the clique indicator M as follows.

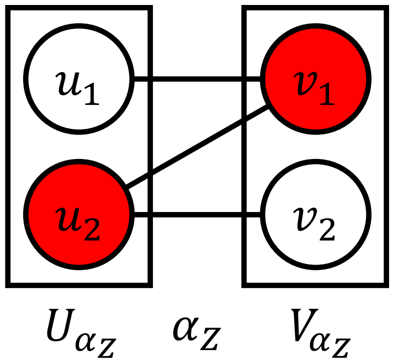
$$M = \frac{1}{2^6} \sum_{\substack{\alpha: U_\alpha = (u_1,u_2), \\ V_\alpha = (v_1,v_2), \\ V(\alpha) = U_\alpha \cup V_\alpha}} M_\alpha$$



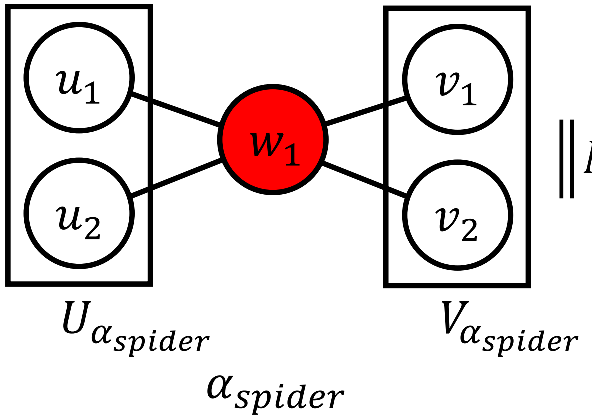
- Idea: If $A \cup B$ is a 4-clique then for all of these shapes α, $M_\alpha(A,B) = 1$. If $A \cup B$ is missing an edge then there is perfect cancellation between the shapes α which have the corresponding edge and the shapes which do not.

Graph Matrix Norm Bounds

- Theorem [AMP20]: For all shapes α with no isolated vertices outside of $U_\alpha \cup V_\alpha$, letting S_α be a minimum vertex separator between U_α and V_α, with high probability $\|M_\alpha\|$ is $\tilde{O}(n^{\frac{|V(\alpha)|-|S_\alpha|}{2}})$.
- Examples: With high probability,



$\|M_{\alpha_Z}\|$ is $\tilde{O}(n^{\frac{4-2}{2}}) = \tilde{O}(n)$



$\|M_{\alpha_{spider}}\|$ is $\tilde{O}(n^{\frac{5-1}{2}}) = \tilde{O}(n^2)$

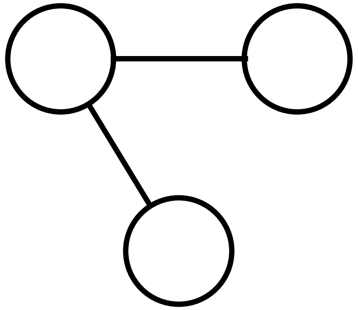
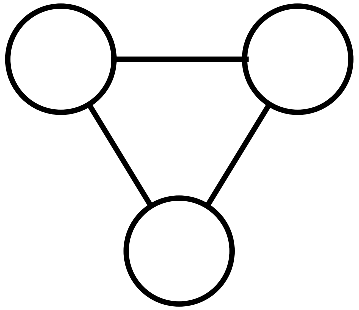
One minimum vertex separator is shown in red.

Pseudo-calibration and Graph Matrices

- Graph matrices are a natural way to represent the moment matrix M given by pseudo-calibration.
- Recall: For planted clique, $\tilde{E}[x_V] = \sum_{E:|V \cup V(E)| \leq t} \left(\frac{k}{n}\right)^{|V \cup V(E)|} \chi_E$
- Decomposition of the moment matrix M using graph matrices:

  $M = \sum_{\alpha:|E(\alpha)| \leq t} \left(\frac{k}{n}\right)^{|V(\alpha)|} M_\alpha.$
- $\tilde{E}[1] = 1 + \sum_{\alpha:U_\alpha = V_\alpha = \emptyset,\, 0 < |E(\alpha)| \leq t} \left(\frac{k}{n}\right)^{|V(\alpha)|} M_\alpha.$

Low-Degree Polynomial Lower Bound Picture

$\tilde{E}[1] = 1 + \left(\frac{k}{n}\right)^2$  $+ \left(\frac{k}{n}\right)^3$  $+ \left(\frac{k}{n}\right)^3$  $+ \cdots$

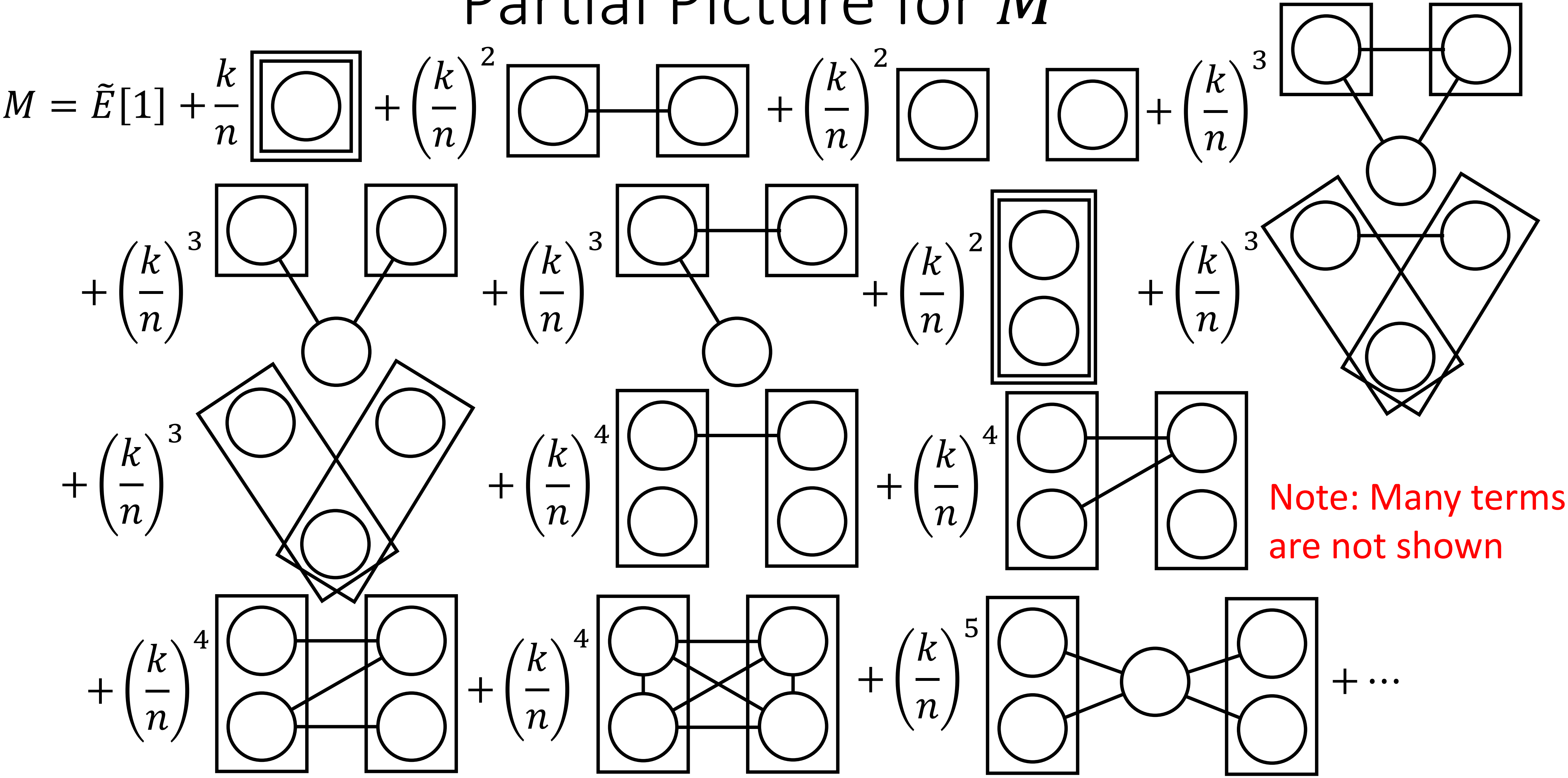
Rough analysis using graph matrices: For all $j \in [2d]$, there are at most $2^{j^2/2}$ shapes α such that $|V(\alpha)| = j$ and $U_\alpha = V_\alpha = \emptyset$. With high probability, all of these terms have magnitude $\tilde{O}(n^{j/2})$.

Using a union bound, we obtain that with high probability,

$$\left|\tilde{E}[1] - 1\right| \leq \sum_{j=1}^{2d} \tilde{O}\left(\left(\frac{2^d k}{\sqrt{n}}\right)^j\right).$$

which is $o(1)$ if $k \ll \sqrt{n}$

Partial Picture for M



Part VII: Current Sum of Squares Lower Bounds for Average Case Problems

Evidence for the Low-Degree Conjecture

- We have SoS lower bounds matching (up to lower order terms) the best known low-degree polynomial lower bounds for
  - Planted clique [BHKKMP16].
  - Random CSPs [KMOW17].
  - Tensor PCA (principal component analysis) and sparse PCA [HKPRSS17, PR20]
  - k-Coloring [KM21]
  - Densest k-subgraph [JPRX23].
  - Non-Gaussian Component Analysis [DKPP24] (SoS lower bounds for a special case were shown in [GJJPR20]).
- For independent set on sparse random graphs (i.e., $G(n,p)$ where p is small), the distinguishing problem is easy but there are SoS lower bounds for certifying that $G(n,p)$ does not have a large independent set [JPRTX21, KPX24] and low-degree polynomial lower bounds for recovering the independent set [SW22].

Potential Improvements

- While we have made quite a bit of progress in understanding the performance of SoS on average case problems, there is still room for improvement. Some potential improvements are as follows.
  1. The current machinery for SoS lower bounds has trouble handling global constraints. For example, the SoS lower bound for planted clique [BHKKMP16] does not satisfy the constraint that the clique has size exactly k. While Shuo Pang [Pang21] resolved this issue for planted clique, we currently don't have general techniques for handling global constraints.
  2. The current machinery for SoS lower bounds relies on the random input being a product distribution. We would like to have techniques for handling other random inputs such as random d-regular graphs.
  3. For robust estimation problems, we often have indicators for whether a sample is corrupted. Our SoS lower bound for NGCA does not include this kind of indicator.

Potential Future SoS Lower Bounds for Average Case Problems

- Currently, the SoS lower bounds for k-coloring [KM21] allows each vertex to have multiple colors. We would like to prove an SoS lower bound for k-coloring where each vertex can only have one color.
- Recently, low-degree lower bounds have been proved for distinguishing between two planted distributions.
  - For low-degree polynomials, counting the number of planted communities in a graph is as hard as recovering the communities [RSWY23].
  - When $n^{3/2} \ll k \ll n$^2, it is hard for low-degree polynomials to distinguish between an order 3 tensor of rank k with random components where all components have coefficient 1 and an order 3 tensor of rank k with random components where the first component has coefficient $1 + \delta$ and the remaining components have coefficient 1 [Wein23].
- Proving SoS lower bounds for distinguishing between two planted distributions would be very interesting.

Some Open Problems

- Can we prove an SoS version of the low-degree conjecture or find natural average-case problems where SoS is significantly stronger than low-degree polynomials?
- Can we strengthen the machinery for proving SoS lower bounds to handle global constraints, non-product input distributions such as $G(n,p)$, and/or indicator variables for whether we take samples?
- Can we prove SoS lower bounds for distinguishing between two planted distributions?
- Can we find a quiet planting for independent set on sparse random graphs?
- Can we prove an SoS lower bound for k-coloring where each vertex has exactly one color?

Thank You!

References

- [AMP20] K. Ahn, D. Medarametla, and A. Potechin. Graph Matrices: Norm Bounds and Applications. arXiv 1604.03423, 2020
- [BHKX22] M. Bafna, J. T. Hsieh, P. Kothari, and J. Xu. Polynomial-Time Power-Sum Decomposition of Polynomials. FOCS 2022
- [BHKKMP16] B. Barak, S. Hopkins, J. Kelner, P. Kothari, A. Moitra, and A. Potechin. A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem. SIAM Journal on Computing Vol. 48, Issue 2, p.687-735, 2019
- [CP20] W. Cai and A. Potechin. The Spectrum of the Singular Values of Z-Shaped Graph Matrices. arXiv 2006.14144, 2020
- [CP22] W. Cai and A. Potechin. On Mixing Distributions Via Random Orthogonal Matrices and the Spectrum of the Singular Values of Multi-Z Shaped Graph Matrices. arXiv 2206.02224, 2022
- [GJJPR20] M. Ghosh, F. G. Jeronimo, C. Jones, A. Potechin, and G. Rajendran. Sum-of-Squares Lower Bounds for Sherrington-Kirkpatrick via Planted Affine Planes. FOCS 2020[Gri01a]
- [Gri01a] D. Grigoriev. Complexity of Positivstellensatz proofs for the knapsack. Computational Complexity 10(2), p. 139–154. 2001
- [Gri01b] D. Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. Theor. Comput. Sci., 259(1-2), p. 613–622. 2001

References

- [HW21] J. Holmgren and A. S. Wein. Counterexamples to the Low-Degree Conjecture. ITCS 2021
- [Hop18] S. B. Hopkins. Statistical Inference and the Sum of Squares Method. PhD thesis, Cornell University, 2018
- [HKPRSS17] S. Hopkins, P. Kothari, A. Potechin. P. Raghavendra, T. Schramm, and D. Steurer. The Power of Sum-of-Squares for Detecting Hidden Structures. FOCS 2017
- [HKPX23] J. Hsieh, P. Kothari, A. Potechin, and J. Xu. Ellipsoid Fitting Up to a Constant. ICALP 2023
- [JP24] C. Jones and L. Pesenti. Diagram analysis of iterative algorithms. arXiv 2404.07881. 2024
- [JPRTX21] C. Jones, A. Potechin, G. Rajendran, M. Tulsiani, J. Xu. Sum-of-Squares Lower Bounds for Sparse Independent Set. FOCS 2021
- [JPRX23] C. Jones, A. Potechin, G. Rajendran, J. Xu. Sum-of-Squares Lower Bounds for Densest k-Subgraph. STOC 2023
- [KM21] P. Kothari and P. Manohar. A Stress-Free Sum-Of-Squares Lower Bound for Coloring. CCC 2021
- [KMOW17] P. Kothari, R. Mori, R. O'Donnell, and D. Witmer. Sum of squares lower bounds for refuting any CSP. STOC 2017
- [KPX24] P. Kothari, A. Potechin, and J. Xu. Sum-of-Squares Lower Bounds for Independent Set on Ultra-Sparse Random Graphs. STOC 2024

References

- [KWB22] D. Kunisky, A. S. Wein, and A. S. Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. ISAAC 2022
- [Pang21] S. Pang. SOS lower bound for exact planted clique. CCC 2021
- [PR20] A. Potechin and G. Rajendran. Machinery for Proving Sum-of-Squares Lower Bounds on Certification Problems. arXiv 2011.04253, 2020. Note: A version of this paper appeared in NeurIPS 2022 with the title “Sub-exponential time Sum-of-Squares lower bounds for Principal Components Analysis”
- [PTVW23] A. Potechin, P. Turner, P. Venkat, and A. Wein. Near-optimal fitting of ellipsoids to random points. COLT 2023
- [RT23] G. Rajendran and M. Tulsiani. Concentration of polynomial random matrices via Efron-Stein inequalities. SODA 2023
- [RSWY23] C. Rush, F. Skerman, A. Wein, and D. Yang. Is it easier to count communities than find them? ITCS 2023
- [SW22] T. Schramm and A. S. Wein. Computational Barriers to Estimation from Low-Degree Polynomials. The Annals of Statistics, Vol. 50, Issue 3, p.1833-1858. 2022
- [Wein23] A. S. Wein. Average-Case Complexity of Tensor Decomposition for Low-Degree Polynomials. STOC 2023

Appendix: 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, $x^T Mx \leq \lambda_{max}(M)\|x\|^2$.
- 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 $Q \preccurlyeq 0$.
- [HKPRSS17]: 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.

Potential 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 $q = O(logn)$.