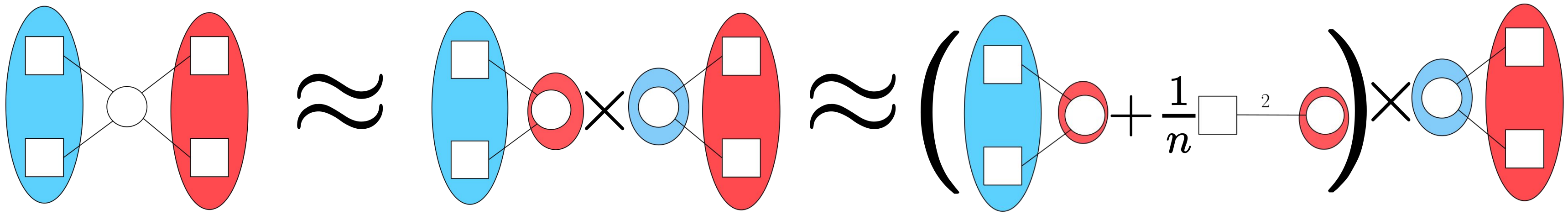
Sum-of-Squares Lower Bounds for Sherrington-Kirkpatrick via Planted Affine Planes



Mrinalkanti Ghosh - TTIC
Fernando Granha Jeronimo - University of Chicago
Chris Jones - University of Chicago
Aaron Potechin - University of Chicago
Goutham Rajendran - University of Chicago

Introduction

Sherrington-Kirkpatrick (SK) problem: Given $W \in \mathbb{R}^{n \times n}$ sampled from GOE(n), find the maximum value of $x^T W x$ over $x \in \{+1, -1\}^n$

GOE(n): Gaussian Orthogonal Ensemble. A symmetric n x n matrix where the off-diagonal entries are N(0, 1) and the diagonal entries are N(0, 2).

Important problem considered before in

- Computer Science - Can be viewed as average-case MaxCut on random graphs where x encodes the partition
- Statistical Physics - Can be viewed as minimizing energy of a physical system, where x encodes spin values in a spin-glass model

Facts about SK

Define OPT(W) = max $(x^T W x)$ over all $x \in \{+1, -1\}^n$

OPT(W) $\leq (2 + o(1))\, n^{1.5}$ whp over W ~ GOE(n), because W has maximum eigenvalue $(2 + o(1))\, n^{0.5}$ whp

[Conjectured by Parisi ’79, proved by Talagrand ’06]
OPT(W) $\approx 2\, P^*\, n^{1.5} \approx 1.52\, n^{1.5}$ whp over W ~ GOE(n)

Here, P* is known as the Parisi constant

Optimization vs Certification

Optimization problem: Given $W \sim GOE(n)$, maximize $x^T W x$ over $x \in \{+1, -1\}^n$

Recently, a breakthrough result by Montanari provided a polynomial time optimization algorithm for SK!

[Montanari '19] There is a poly(n, 1/ε) time algorithm that w.h.p. outputs $x \in \{+1, -1\}^n$ satisfying $x^T W x \geq (2P^* - \varepsilon)\, n^{1.5}$ *(under a statistical physics conjecture)*

In this work, we are concerned with the certification problem:
Given $W \sim GOE(n)$, output an upper bound on max $(x^T W x)$ over $x \in \{+1, -1\}^n$ that gets close to the true optimum whp.

The Sum-of-Squares hierarchy

Sum-of-Squares is a powerful certification method for a polynomial objective/constraints.

The algorithm is parameterized by degree d, with larger d = more powerful SoS.

Main question: what bound can SoS certify for the SK problem?

[Montanari-Sen '16, Kunisky-Bandeira '19, Mohanty-Raghavendra-Xu '20, Kunisky '20]
Whp degree-6 SoS has value $(2 - o(1))\, n^{1.5}$

This work: For some constant $\delta > 0$, whp degree-n^{δ} SoS still has value $(2\text{-}o(1))\, n^{1.5}$

Rules out degree-O(1) SoS = polynomial time SoS

Planted Affine Planes

Planted Affine Planes (PAP) problem: Let $n \ll m$. If $d_1, \ldots, d_m$ are random vectors in R^n sampled from $N(0, I_n)$, can we prove that there is no vector $v \in R^n$ such that $\langle v, d_u \rangle^2 = 1$ for all $u = 1, 2, \ldots, m$?

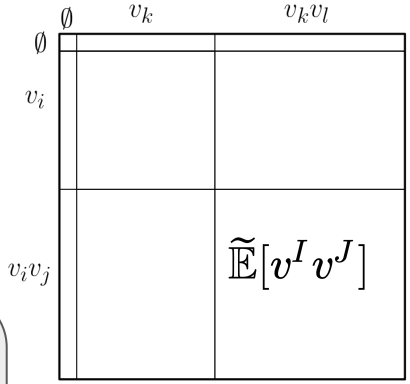
This work: For $m \le n^{1.5 - \varepsilon}$, w.h.p. over $d_1, \ldots, d_m \sim N(0, I_n)$, degree-$n^\delta$ SoS thinks this system of equations is feasible.

Remainder of talk: SoS lower bound for PAP

[BHKKMP '16] Recipe for SoS lower bounds

Goal for SoS lower bounds: Construct degree-D pseudodistribution of solutions, specified by pseudomoments $\widetilde{\mathbb{E}}[v^S]$ for all subsets S of {1, ..., n} of size at most D.

Equivalently, construct a moment matrix $\mathcal{M}$ with rows, columns indexed by subsets of {1, .., n} of size ≤ D/2 such that (1) $\mathcal{M}$ obeys some linear constraints on its entries, and (2) $\mathcal{M} \succcurlyeq 0$.

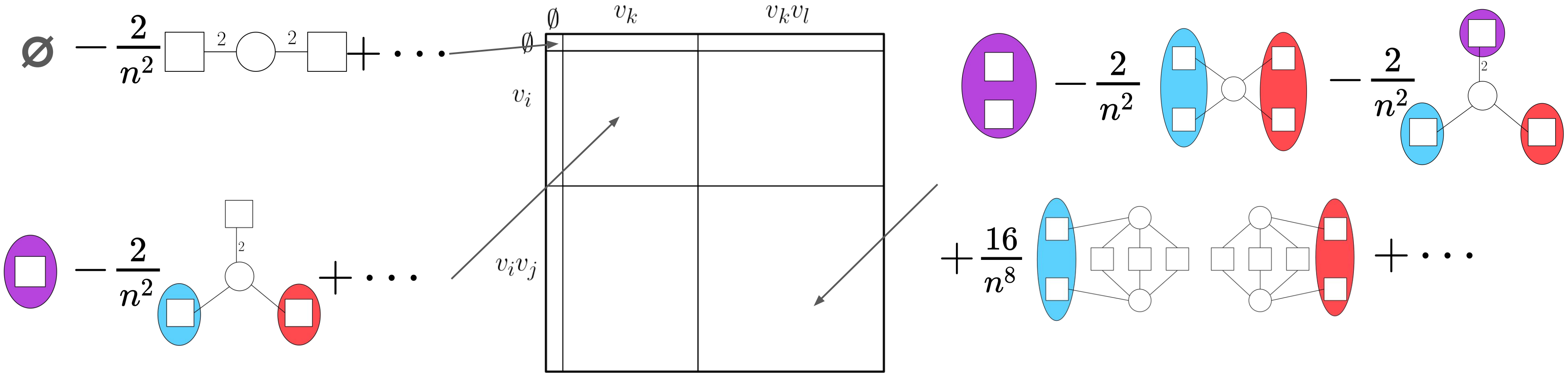


[BHKKMP '16] Recipe for average-case problems

1. Construct a candidate $\mathcal{M}$ via pseudocalibration
2. Decompose $\mathcal{M}$ into graph matrices and use the decomposition to prove $\mathcal{M} \succcurlyeq 0$

Graph matrix decomposition

$$\mathcal{M} = \sum_{\text{shapes } \beta} c_\beta M_\beta = \sum_{\substack{\text{shapes } \beta: \\ \beta \text{ satisfies parity constraints}}} \prod_{\circ \in V(\beta)} h_{\deg(\circ)}(1) \cdot \frac{M_\beta}{n^{|E(\beta)|/2}}$$



Theorem [AMP '20]: Whp for all shapes β: let S be a minimum-weight vertex separator of the left and right sides of β,

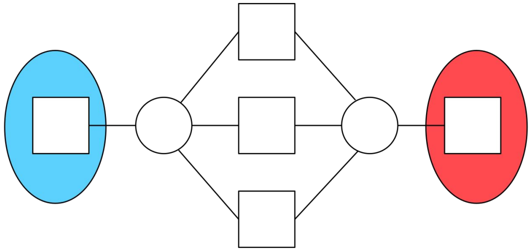
$$\|M_\beta\| \le \tilde{O}\left(\sqrt{m}^{\#\circ \text{ not in S}} \cdot \sqrt{n}^{\#\square \text{ not in S}}\right)$$

Graph matrix decomposition

$$\mathcal{M} = \sum_{\text{shapes } \beta} c_\beta M_\beta = \sum_{\substack{\text{shapes } \beta: \\ \beta \text{ satisfies parity constraints}}} \prod_{\circ \in V(\beta)} h_{\deg(\circ)}(1) \cdot \frac{M_\beta}{n^{|E(\beta)|/2}}$$

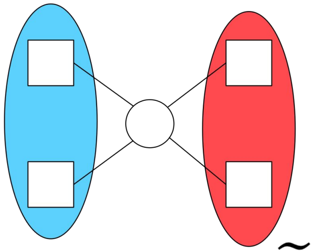
Naive idea: show that $\sum_{\beta \neq \text{identity}} |c_\beta| \|M_\beta\| = o(1)$

Some terms do have o(1) norm...



$$|c_\beta| \cdot \|M_\beta\| = o(1)$$

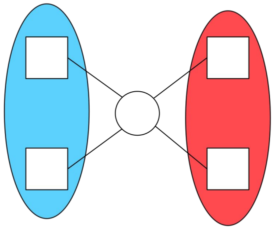
....but others do not

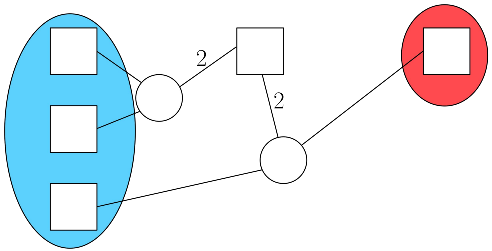


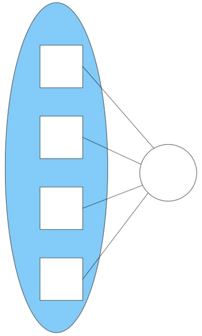
$$|c_\beta| \cdot \|M_\beta\| = \widetilde{\Omega}(1)$$

Spiders

Def: a spider has two degree-1 squares in left side or right side adjacent to same circle







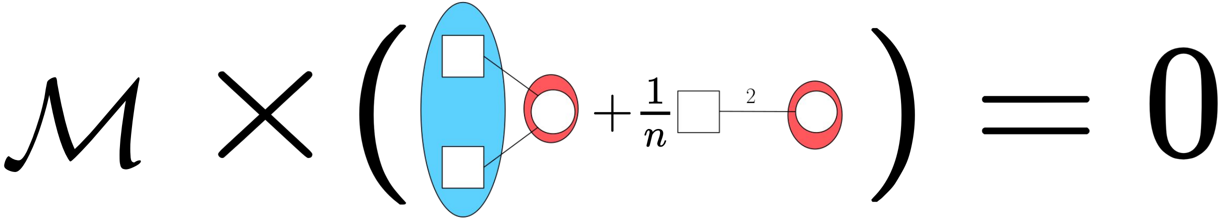
Theorem: $$\sum_{\substack{\text{non-spider,}\\ \text{non-identity } \beta}} |c_\beta| \cdot \|M_\beta\| = o(1)$$

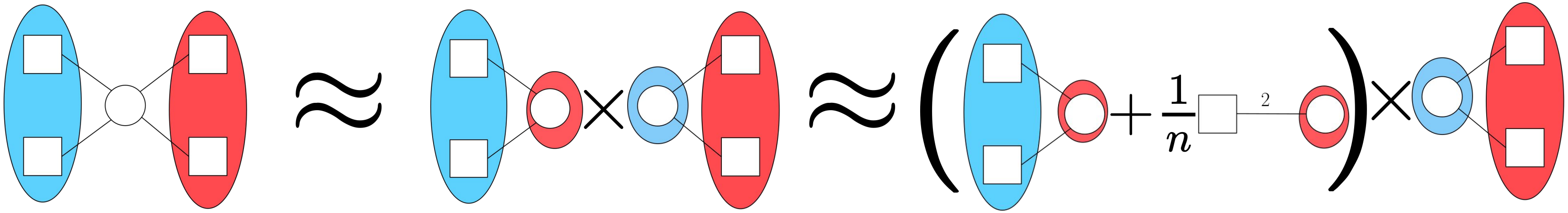
So the only large-norm shapes are the spiders.

Spiders

It suffices to prove that $\mathcal{M}$ is PSD on Null($\mathcal{M}$)$^\perp$

Spiders (the large-norm shapes) are approximately zero on Null($\mathcal{M}$)$^\perp$!





Theorem: There is $\mathcal{M}'$ such that $x^T \mathcal{M} x = x^T \mathcal{M}' x$ for all $x \in$ Null($\mathcal{M}$)$^\perp$ and all eigenvalues of $\mathcal{M}'$ are $1 \pm o(1)$

Open Problems

Interpret Montanari's algorithm as rounding SoS?

Average-case sparse MaxCut

Improve PAP assumption $m \le n^{1.5}$ to $m \le n^2$

Improve SoS degree to $n/\log n$

Proof outline

Reduce the SK problem to the Planted Boolean Vector problem [Mohanty-Raghavendra-Xu ‘19]

Consider the dual version of Planted Boolean Vector problem, which we term Planted Affine Planes (PAP)

Directly prove an SoS lower bound for PAP via the following steps

1. Construct a candidate solution using pseudocalibration
2. Prove positive-semidefiniteness of the candidate moment matrix M

The PSDness proof is the most innovative part of this work

Outline

1. The Sherrington-Kirkpatrick problem
2. Reduction to Planted Affine Planes
3. Pseudocalibration + graph matrices
4. PSD-ness sketch
5. Open Problems

The Sum-of-Squares hierarchy

Sum-of-Squares looks for a refutation proof of a system of polynomial constraints.

A degree-D pseudodistribution of solutions exists iff no degree-D SoS refutation exists
In other words, SoS thinks the system is feasible.

Dual view: A series of convex relaxations to a polynomial program, parameterized by D, called degree of SoS.

Obtains state-of-the-art algorithms for many problems such as Max k-CSPs, Tensor PCA, etc.

Question restated: How do SoS relaxations for the Sherrington-Kirkpatrick problem perform?

SoS lower bounds for PAP

Main theorem restated: For $m \le n^{1.5 - \varepsilon}$, w.h.p. over $d_1, \dots, d_m \sim N(0, I_n)$, degree-$n^\delta$ SoS thinks this system of equations is feasible.

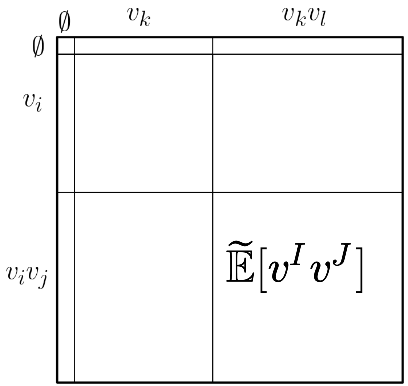
More concretely, given input $d_1, \dots, d_m$ in R^n, specify $\widetilde{\mathbb{E}}[v_1]$, $\widetilde{\mathbb{E}}[v_2 v_4]$, etc., such that

1. $\widetilde{\mathbb{E}}[1] = 1$ and $\widetilde{\mathbb{E}}[\langle v, d_u \rangle^2] = \widetilde{\mathbb{E}}[1]$ for all u = 1, ..., m and other constraints.
2. $\widetilde{\mathbb{E}}[g^2] \ge 0$ for all polynomials of degree at most D/2. **[Positivity condition]**

Moment matrix: Matrix $\mathcal{M}$ with rows, columns indexed by subsets of {1, .., n} such that

$\mathcal{M}[I, J] = \widetilde{\mathbb{E}}[v^I \cdot v^J]$ for subsets I, J of {1, .., n} of size ≤ D/2

Positivity condition: $\mathcal{M} \succcurlyeq 0$



Pseudocalibration

The pseudocalibration heuristic introduced by [BHKKMP '16] gives a candidate $\widetilde{\mathbb{E}}$ that satisfies the constraints (approximately).

For any subset S ⊆ [n], $\widetilde{\mathbb{E}}[v^S] = \sum_{\alpha \in \mathbb{N}^{m \times n}} c_\alpha h_\alpha(d_1, \ldots, d_m)$

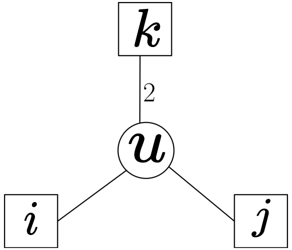
h_α - basis of Hermite polynomials
c_α - real coefficients
c_α = 0 unless α satisfies simple parity conditions, in which case, $c_\alpha \approx n^{-|\alpha|/2}$

Main difficulty: Checking the positivity condition $\mathcal{M} \succcurlyeq 0$

Graph Matrices

Hermite polynomials on $\{d_u\}$ are in 1-to-1 correspondence with $\mathbb{N}$-edge-labeled graphs on $[m] \cup [n]$

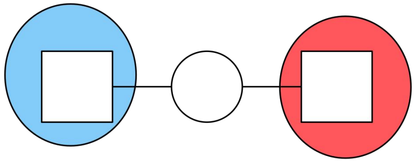
$$h_1(d_{u,i})h_1(d_{u,j})h_2(d_{u,k}) \longleftrightarrow$$



If we look at $\mathcal{M}$, the same "Fourier shape" appears in lots of different entries $\mathcal{M}$[I, J]. And in fact the coefficient c_α only depends on (1) the shape and (2) the sets I, J

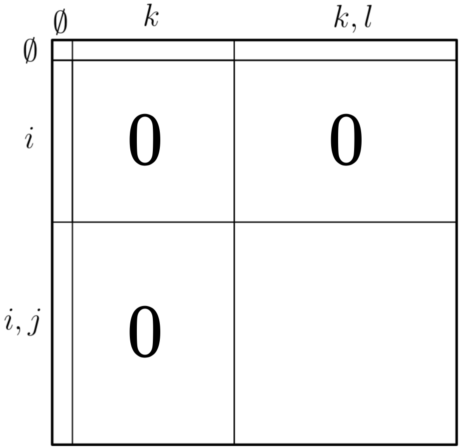
Collect all such entries together into a matrix M_β encoded by a graph β

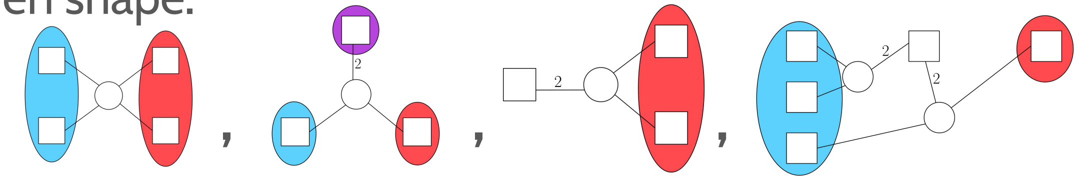
This will be a graph matrix.

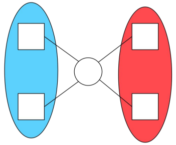


Graph Matrices

Graph matrix [BHKKMP '16, AMP '20]: A graph matrix M_β is defined by a shape β = bipartite ℕ-edge-labeled graph with special subsets of vertices U, V. The matrix sums up all Hermite characters with the given shape.





For β =  ,

$$M_\beta[\{i,j\},\{k,l\}] = \sum_{u=1}^{m} h_1(d_{u,i})h_1(d_{u,j})h_1(d_{u,k})h_1(d_{u,l})$$

$$\mathcal{M} = \sum_{\substack{\text{shapes } \beta: \\ \beta \text{ satisfies parities constraints}}} \prod_{\circ \in V(\beta)} h_{\deg(\circ)}(1) \cdot \frac{M_\beta}{n^{|E(\beta)|/2}}$$

Graph Matrices

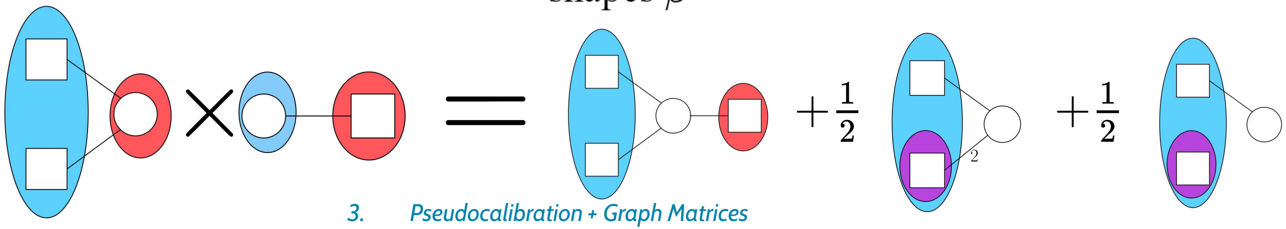
Theorem [AMP '20]: W.h.p. for all shapes β: let S be a minimum-weight vertex separator of left and right sides of β,

$$\|M_\beta\| \leq \widetilde{O}\left(\sqrt{m}^{\#\bigcirc \text{ not in S}} \cdot \sqrt{n}^{\#\square \text{ not in S}}\right)$$

In fact, this is the only property of the random input that we need

Graph matrices are a "functional matrix algebra":

$$M_\beta \cdot M_\gamma = \sum_{\text{shapes } \beta'} c_{\beta'} M_{\beta'}$$



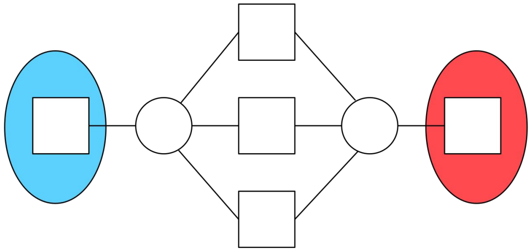
PSD-ness proof sketch

Norm bounds:
$\|M_\beta\| \leq$
$\widetilde{O}(\sqrt{m}^{\#\circ \text{ not in S}} \cdot \sqrt{n}^{\# \square \text{ not in S}})$
where S is a min-weight vertex separator of β

$$\mathcal{M} = \sum_{\text{shapes } \beta} c_\beta M_\beta = \sum_{\substack{\text{shapes } \beta: \\ \beta \text{ satisfies parity constraints}}} \prod_{\circ \in V(\beta)} h_{\deg(\circ)}(1) \cdot \frac{M_\beta}{n^{|E(\beta)|/2}}$$

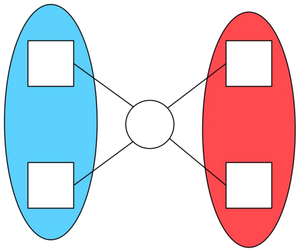
Naive idea: show that $\sum_{\beta \neq \text{identity}} |c_\beta| \|M_\beta\| = o(1)$

Some terms do have o(1) norm...



$|c_\beta| \cdot \|M_\beta\| = O(1/n^4) \cdot \widetilde{O}(n^2 m)$
$= \widetilde{O}(m/n^2) = o(1)$

....but others do not

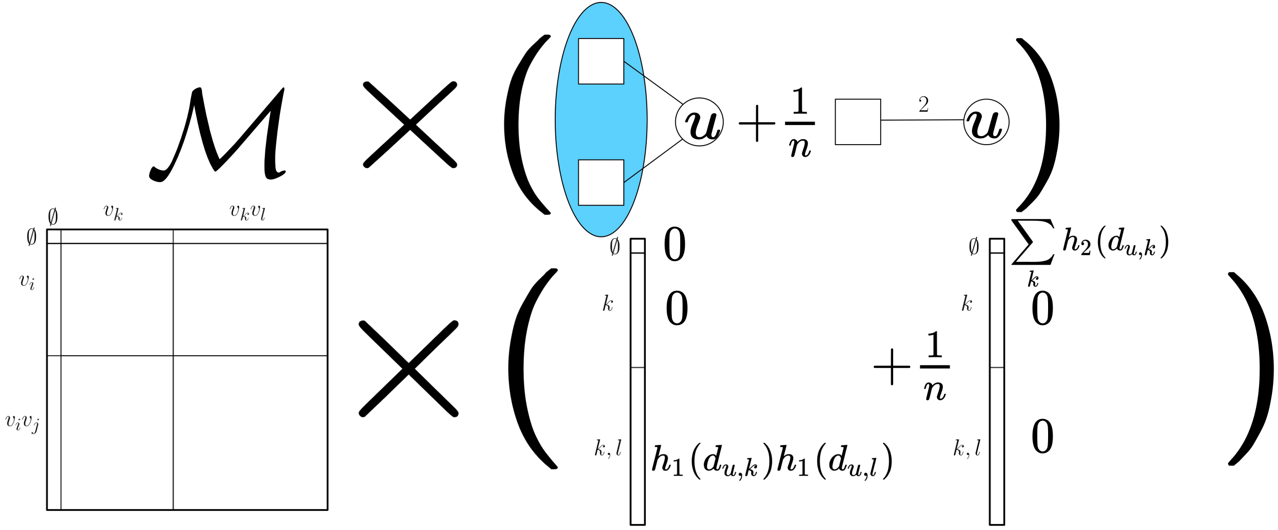


$|c_\beta| \cdot \|M_\beta\| = O(1/n^2) \cdot \widetilde{O}(n^2)$
$= \widetilde{\Omega}(1)$

PSD-ness proof sketch

Non-identity terms with Ω(1) norm *must* exist because $\mathcal{M}$ has a nontrivial null space

Null space is induced by constraints "$\langle v, d_u \rangle^2 = 1$"



After some simplifications, top row $= \widetilde{\mathbb{E}}[\langle v, d_u \rangle^2 - 1]$

Certification

Certification problem: Given $W \sim GOE(n)$, output an upper bound on max $(x^T W x)$ over $x \in \{+1, -1\}^n$ that gets close to the true optimum whp.

Spectral certificate: Just output the maximum eigenvalue of W, which is $(2 + o(1))\, n^{1.5}$ whp.

Can other certificates get closer to the true optimum, which is $\approx 1.52\, n^{1.5}$ whp?

In particular, the Sum-of-Squares hierarchy offers a natural series of convex relaxations for this problem.

Results on SoS for SK

Degree-2 SoS has value $(2 - o(1))\, n^{1.5}$
[Montanari-Sen '16]

Degree-4 SoS has value $(2 - o(1))n^{1.5}$
[Mohanty-Raghavendra-Xu '20; Kunisky-Bandeira '19]

Degree-6 SoS has value $(2 - o(1))\, n^{1.5}$
[Kunisky '20]

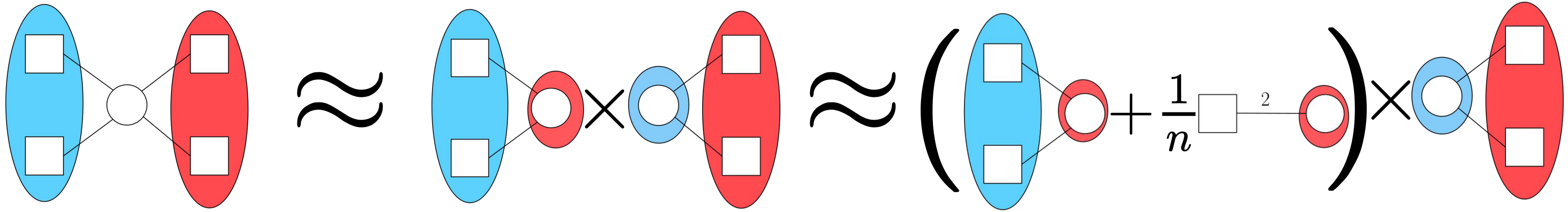
This work: For some constant $\delta > 0$, whp degree-n^{δ} SoS still has value $(2\text{-}o(1))\, n^{1.5}$

Rules out degree-$O(1)$ SoS = polynomial time SoS

PSD-ness proof sketch

It suffices to prove that $\mathcal{M}$ is PSD on Null($\mathcal{M}$)$^\perp$

Spiders approximately factor into a matrix with columns from Null($\mathcal{M}$)!



Thus spiders are approximately zero on Null($\mathcal{M}$)$^\perp$

$$x^T \mathcal{M} x \approx x^T (\mathcal{M} - \text{spider})\, x \text{ for all } x \in \text{Null}(\mathcal{M})^\perp$$

Theorem: There is $\mathcal{M}'$ such that $x^T \mathcal{M} x = x^T \mathcal{M}' x$ for all $x \in$ Null($\mathcal{M}$)$^\perp$ and all eigenvalues of $\mathcal{M}'$ are $1 \pm o(1)$

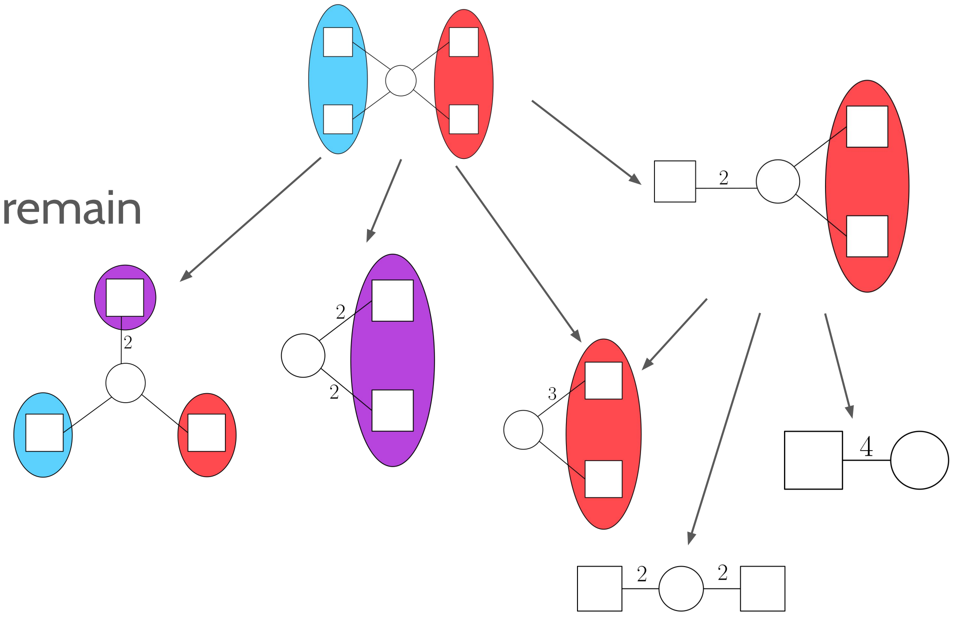
PSD-ness proof sketch

Due to ≈, killing a spider introduces smaller terms into the moment matrix

$$x^T \mathcal{M} x = x^T (\mathcal{M} - \text{spider} + \text{intersection terms})\, x \text{ for all } x \in \text{Null}(\mathcal{M})^\perp$$

Some of these may be smaller spiders!

Recursively kill these until only non-spiders remain



Theorem: For c_β' the new coefficients on non-spiders,

$$\sum_{\substack{\text{non-spider,} \\ \text{non-identity } \beta}} |c'_\beta| \cdot \|M_\beta\| = o(1)$$

Open Problems

Interpret Montanari's algorithm as rounding SoS?

Average case sparse MaxCut

Improve PAP assumption $m \leq n^{1.5}$ to $m \leq n^2$

Improve SoS degree to n/log n