Nonlinear Random Matrices and Applications to the Sum of Squares Hierarchy

Goutham Rajendran

Committee: Aaron Potechin, Janos Simon, Madhur Tulsiani

June 13, 2022

THE UNIVERSITY OF CHICAGO

Table of Contents

Table of Contents

Optimization vs Certification

Optimization problems in computer science seek to optimize an objective. For example,

- Given a graph, what is the largest cut?
- Given a set of points, find the shortest distance between a pair of points

Optimization vs Certification

Optimization problems in computer science seek to optimize an objective. For example,

- Given a graph, what is the largest cut?
- Given a set of points, find the shortest distance between a pair of points

Certification problems ask for both bounds on the objective and a *corresponding certificate*, w.h.p. over the input, e.g.,

- Given a graph, certify an upper bound on the largest cut
- Given a set of points, certify an upper bound on the shortest distance between a pair of points

Optimization vs Certification

Optimization problems in computer science seek to optimize an objective. For example,

- Given a graph, what is the largest cut?
- Given a set of points, find the shortest distance between a pair of points

Certification problems ask for both bounds on the objective and a *corresponding certificate*, w.h.p. over the input, e.g.,

- Given a graph, certify an upper bound on the largest cut
- Given a set of points, certify an upper bound on the shortest distance between a pair of points

We study certification for average-case problems that are "good": Close to optimum w.h.p.

Average-case problems

Motivation

Are hard instances of optimization problems *natural* instances?

Average-case problems

Motivation

Are hard instances of optimization problems *natural* instances?

For example, the simplex algorithm for Linear Programming is exponential time in worst case but works very well in practice

Average-case problems

Motivation

Are hard instances of optimization problems *natural* instances?

For example, the simplex algorithm for Linear Programming is exponential time in worst case but works very well in practice

In this work, we study certification problems on random inputs, e.g.

- Given an Erdős-Rényi random graph $G \sim \mathcal{G}_{n,p}$, certify an upper bound on the size of the maximum independent set
- Given a matrix W sampled from the Gaussian Orthogonal Ensemble, certify an upper bound on $x^T W x$ where x is boolean

Computational complexity of Average-case problems

A certification algorithm $\mathcal{A}$ certifies a bound U for an average-case problem if w.h.p. over the input, $\mathcal{A}$ outputs $U + o(1)$

We study the existence of efficient certification algorithms

Computational complexity of Average-case problems

A certification algorithm $\mathcal{A}$ certifies a bound U for an average-case problem if w.h.p. over the input, $\mathcal{A}$ outputs $U + o(1)$

We study the existence of efficient certification algorithms

For worst-case analysis, NP-hardness is a gold standard for computational hardness

For average-case analysis, NP-hardness results are beyond reach

Computational complexity of Average-case problems

A certification algorithm $\mathcal{A}$ certifies a bound U for an average-case problem if w.h.p. over the input, $\mathcal{A}$ outputs $U + o(1)$

We study the existence of efficient certification algorithms

For worst-case analysis, NP-hardness is a gold standard for computational hardness

For average-case analysis, NP-hardness results are beyond reach

Instead, we study limits of restricted classes of algorithms, e.g.,

- algorithms based on low-degree polynomials
- statistical query algorithms
- the Sum-of-Squares hierarchy of algorithms ⟵ **this work**

Why these? They capture a wide variety of algorithmic techniques, e.g. local reasoning and spectral methods

Main theme of this work: Nonlinear random matrices

When analyzing SoS on average case problems, the main difficulty comes down to analyzing random matrices

Example: What is the maximum eigenvalue of the adjacency matrix of a graph $G \sim \mathcal{G}_{n,1/2}$?

Main theme of this work: Nonlinear random matrices

When analyzing SoS on average case problems, the main difficulty comes down to analyzing random matrices

Example: What is the maximum eigenvalue of the adjacency matrix of a graph $G \sim \mathcal{G}_{n,1/2}$?

Random matrix theory has been studied since 1900s, with applications in mathematics, physics and computer science

Main theme of this work: Nonlinear random matrices

When analyzing SoS on average case problems, the main difficulty comes down to analyzing random matrices

Example: What is the maximum eigenvalue of the adjacency matrix of a graph $G \sim \mathcal{G}_{n,1/2}$?

Random matrix theory has been studied since 1900s, with applications in mathematics, physics and computer science

In many TCS applications, the kind of matrices that appear are nonlinear
Example: Encode graph G as $G_{ij} \in \{-1, 1\}$. Consider

$$\begin{array}{ccc} & \text{column } (k) & \\ & \downarrow & \\ \text{row } (i,j) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & G_{ij} G_{jk} G_{ik} & \cdots \\ & \vdots & \end{pmatrix} & O(n^2) \text{ rows} \\ & n \text{ columns} & \end{array}$$

A summary of our contributions

A general concentration inequality for nonlinear random matrices based on Matrix Efron-Stein inequalities [1]

[1] [**R**, Tulsiani, 2021] - In submission

UChicago affiliation is highlighted in blue

A summary of our contributions

A general concentration inequality for nonlinear random matrices based on Matrix Efron-Stein inequalities [1]

SoS lower bounds

Problem	Informal statement
Sherrington-Kirkpatrick Hamiltonian [2]	Given $W \sim GOE(n)$, find $\max_{x \in \{+1,-1\}^n} x^\intercal W x$
Tensor PCA [3]	For random B, given $\lambda u^{\otimes k} + B$, recover the spike u
Sparse PCA [3]	Given $v_1, \ldots, v_m \sim \mathcal{N}(0, I + \lambda v v^\intercal)$ where v is sparse, recover v
Planted Slightly Denser Subgraph [3]	Given $G \sim \mathcal{G}_{n,\frac{1}{2}}$ with a planted subgraph $H \sim \mathcal{G}_{k,p}, p > \frac{1}{2}$, recover it

[1] [**R**, Tulsiani, 2021] - In submission

[2] [Ghosh, Jeronimo, Jones, Potechin, **R**, 2020] - FOCS 2020

[3] [Potechin, **R**, 2020] - In submission

UChicago affiliation is highlighted in blue

Followup and other works

SoS lower bounds for Sparse Independent Set

- First SoS lower bound on sparse Erdős-Rényi graphs
- [Jones, Potechin, **R**, Tulsiani, Xu, 2021] - FOCS 2021

UChicago affiliation is highlighted in blue
CMU affiliation is highlighted in red

Followup and other works

SoS lower bounds for Sparse Independent Set

- First SoS lower bound on sparse Erdős-Rényi graphs
- [Jones, Potechin, **R**, Tulsiani, Xu, 2021] - FOCS 2021

Causal inference and latent variable modeling

- Structure learning in polynomial time - [**R**, Kivva, Gao, Aragam, 2021] - NeurIPS 2021
- Learning latent causal graphs via mixture oracles - [Kivva, **R**, Ravikumar, Aragam, 2021] - NeurIPS 2021

UChicago affiliation is highlighted in blue
CMU affiliation is highlighted in red

Table of Contents

Concentration behavior of random matrices

Let $G \sim \mathcal{G}_{n,1/2}$ be an Erdős-Rényi random graph
Encode it with variables $G_{ij} \in \{-1, 1\}$ for $1 \leq i, j \leq n$

Concentration behavior of random matrices

Let $G \sim \mathcal{G}_{n,1/2}$ be an Erdős-Rényi random graph
Encode it with variables $G_{ij} \in \{-1, 1\}$ for $1 \leq i, j \leq n$

Consider adjacency matrix $$A = \begin{array}{ccc} & \text{column } (j) & \\ & \downarrow & \\ \text{row } (i) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & G_{ij} & \cdots \\ & \vdots & \end{pmatrix} & n \text{ rows} \\ & n \text{ columns} & \end{array}$$

Concentration behavior of random matrices

Let $G \sim \mathcal{G}_{n,1/2}$ be an Erdős-Rényi random graph
Encode it with variables $G_{ij} \in \{-1, 1\}$ for $1 \leq i, j \leq n$

Consider adjacency matrix

$$A = \begin{array}{ccc} & \text{column } (j) & \\ & \downarrow & \\ \text{row } (i) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & G_{ij} & \cdots \\ & \vdots & \end{pmatrix} & n \text{ rows} \\ & n \text{ columns} & \end{array}$$

Concentration question: How much does A deviate from $\mathbb{E}[A] = 0$?

Concretely, can we bound $\|A\|$ whp?

Concentration behavior of random matrices

Let $G \sim \mathcal{G}_{n,1/2}$ be an Erdős-Rényi random graph
Encode it with variables $G_{ij} \in \{-1, 1\}$ for $1 \leq i, j \leq n$

Consider adjacency matrix

$$A = \begin{array}{ccc} & \text{column } (j) & \\ & \downarrow & \\ \text{row } (i) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & G_{ij} & \cdots \\ & \vdots & \end{pmatrix} & n \text{ rows} \\ & n \text{ columns} & \end{array}$$

Concentration question: How much does A deviate from $\mathbb{E}[A] = 0$?

Concretely, can we bound $\|A\|$ whp?
Answer: Yes, $O(\sqrt{n})$

Concentration behavior of random matrices

A trickier question: Let $G \sim \mathcal{G}_{n,1/2}$

$$\text{Let } B = \begin{array}{ccc} & \text{column } (k) & \\ & \downarrow & \\ \text{row } (i,j) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & G_{ij}\,G_{jk}\,G_{ik} & \cdots \\ & \vdots & \end{pmatrix} & O(n^2) \text{ rows} \\ & n \text{ columns} & \end{array}$$

Concentration behavior of random matrices

A trickier question: Let $G \sim \mathcal{G}_{n,1/2}$

$$\text{Let } B = \begin{array}{ccc} & \text{column } (k) & \\ & \downarrow & \\ \text{row } (i,j) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots G_{ij} & G_{jk} & G_{ik} \cdots \\ & \vdots & \end{pmatrix} & O(n^2) \text{ rows} \\ & n \text{ columns} & \end{array}$$

Can we find a good bound on $\|B\|$ whp?

Concentration behavior of random matrices

A trickier question: Let $G \sim \mathcal{G}_{n,1/2}$

$$\text{Let } B = \begin{array}{rcl} & \text{column } (k) & \\ & \downarrow & \\ \text{row } (i,j) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & G_{ij}\,G_{jk}\,G_{ik} & \cdots \\ & \vdots & \end{pmatrix} & O(n^2) \text{ rows} \\ & n \text{ columns} & \end{array}$$

Can we find a good bound on $\|B\|$ whp?

We are interested in answering such questions

Polynomial random matrices

Large theory exists to understand the behavior of linear random matrices, e.g, Bernstein's inequality

Far fewer tools available for polynomial random matrices

Polynomial random matrices

Large theory exists to understand the behavior of linear random matrices, e.g, Bernstein's inequality
Far fewer tools available for polynomial random matrices

But such matrices occur often in TCS, e.g., [Barak et al., 2012], [Ge and Ma, 2015], [Hopkins et al., 2015], [Schramm and Steurer, 2017]

Polynomial random matrices

Large theory exists to understand the behavior of linear random matrices, e.g, Bernstein's inequality
Far fewer tools available for polynomial random matrices

But such matrices occur often in TCS, e.g., [Barak et al., 2012], [Ge and Ma, 2015], [Hopkins et al., 2015], [Schramm and Steurer, 2017]

[Moitra and Wein, 2019] popularize a general framework called tensor networks to design such spectral algorithms

Polynomial random matrices

Large theory exists to understand the behavior of linear random matrices, e.g, Bernstein's inequality
Far fewer tools available for polynomial random matrices

But such matrices occur often in TCS, e.g., [Barak et al., 2012], [Ge and Ma, 2015], [Hopkins et al., 2015], [Schramm and Steurer, 2017]

[Moitra and Wein, 2019] popularize a general framework called tensor networks to design such spectral algorithms

A concrete example: To study algorithms for PCA, [Hopkins et al., 2015] bound $\|M - \mathbb{E}[M]\|$ where

$$M = A_1 \otimes A_1 + \ldots + A_m \otimes A_m$$

- Entries are A_i are iid in $\{-1, 1\}$ uniformly at random
- M is a degree 2 polynomial in these variables

Nonlinear random matrices

A widely used method to analyze concentration is trace power method

Nonlinear random matrices

A widely used method to analyze concentration is trace power method

Trace power method

- For a large enough t, bound $\mathbb{E}[\mathrm{tr}[(MM^\intercal)^t]]$
- Appeal to Markov's inequality to get a high probability bound on $\|M\|$

Nonlinear random matrices

A widely used method to analyze concentration is trace power method

Trace power method

- For a large enough t, bound $\mathbb{E}[\text{tr}[(MM^\intercal)^t]]$
- Appeal to Markov's inequality to get a high probability bound on $\|M\|$

Here, $\mathbb{E}[\text{tr}[(MM^\intercal)^t]] = \mathbb{E}\,\|M\|_{2t}^{2t}$ is the expected $2t$-th power of the Schatten-$2t$ norm

If $\lambda_1 \geq \ldots \geq \lambda_n$ are the singular values of M, then

$$n\lambda_1^{2t} \quad \geq \quad \|M\|_{2t}^{2t} = \lambda_1^{2t} + \ldots + \lambda_n^{2t} \quad \geq \quad \lambda_1^{2t}$$

So studying $\|M\| = \lambda_1$ is qualitatively the same as studying $\|M\|_{2t}$

Nonlinear random matrices

A widely used method to analyze concentration is trace power method

Trace power method

- For a large enough t, bound $\mathbb{E}[\mathrm{tr}[(MM^\intercal)^t]]$
- Appeal to Markov's inequality to get a high probability bound on $\|M\|$

Here, $\mathbb{E}[\mathrm{tr}[(MM^\intercal)^t]] = \mathbb{E}\,\|M\|_{2t}^{2t}$ is the expected $2t$-th power of the Schatten-$2t$ norm

If $\lambda_1 \geq \ldots \geq \lambda_n$ are the singular values of M, then

$$n\lambda_1^{2t} \quad \geq \quad \|M\|_{2t}^{2t} = \lambda_1^{2t} + \ldots + \lambda_n^{2t} \quad \geq \quad \lambda_1^{2t}$$

So studying $\|M\| = \lambda_1$ is qualitatively the same as studying $\|M\|_{2t}$

Trace power method gives good bounds but often, it is nontrivial and requires ingenious combinatorics

Motivation: Scalar concentration

In this work, we propose an alternative to trace power method

Motivation: Scalar concentration

In this work, we propose an alternative to trace power method

As motivation, let's consider 1×1 matrices, i.e. scalar polyomials

- For linear concentration, we have Chernoff bounds, Hoeffding's inequalities, etc
- Polynomial concentration is already interesting, e.g. hypercontractivity, Efron-Stein inequalities, works by [Kim and Vu, 2000], [Latała, 2006], [Schudy and Sviridenko, 2011], [Adamczak and Wolff, 2015], etc.

Motivation: Scalar concentration

In this work, we propose an alternative to trace power method

As motivation, let's consider 1×1 matrices, i.e. scalar polyomials

- For linear concentration, we have Chernoff bounds, Hoeffding's inequalities, etc
- Polynomial concentration is already interesting, e.g. hypercontractivity, Efron-Stein inequalities, works by [Kim and Vu, 2000], [Latała, 2006], [Schudy and Sviridenko, 2011], [Adamczak and Wolff, 2015], etc.

We will be particularly interested in Efron-Stein inequalities since it can be generalized for matrices
Let's first look at the scalar version

Scalar Efron-Stein inequality

For independent random variables $Z_1, \ldots, Z_n$, let $Z^{(i)}$ denote $Z_1, \ldots, Z_{i-1}, \widetilde{Z}_i, Z_{i+1}, \ldots, Z_n$, where Z_i has been resampled to $\widetilde{Z}_i$

Scalar Efron-Stein inequality

For independent random variables $Z_1, \ldots, Z_n$, let $Z^{(i)}$ denote $Z_1, \ldots, Z_{i-1}, \widetilde{Z}_i, Z_{i+1}, \ldots, Z_n$, where Z_i has been resampled to $\widetilde{Z}_i$

Scalar Efron-Stein inequality [Boucheron et al., 2005]

For a scalar function $f(Z)$,

$$\mathbb{E}\,(f(Z) - \mathbb{E}\,f)^{2t} \;\leq\; O(t)^t \cdot \mathbb{E}[(V(Z))^t]$$

where $V(Z) \;:=\; \sum_{i \in [n]} \mathbb{E}[(f(Z) - f(Z^{(i)}))^2 | Z]$ is the variance proxy

Efron-Stein inequality bounds the deviation of a function in terms of local variance estimates obtained by changing one variable at a time

Scalar Efron-Stein inequality

For independent random variables $Z_1, \ldots, Z_n$, let $Z^{(i)}$ denote $Z_1, \ldots, Z_{i-1}, \widetilde{Z}_i, Z_{i+1}, \ldots, Z_n$, where Z_i has been resampled to $\widetilde{Z}_i$

Scalar Efron-Stein inequality [Boucheron et al., 2005]

For a scalar function $f(Z)$,

$$\mathbb{E}\,(f(Z) - \mathbb{E}\,f)^{2t} \;\leq\; O(t)^t \cdot \mathbb{E}[(V(Z))^t]$$

where $V(Z) \;:=\; \sum_{i\in[n]} \mathbb{E}[(f(Z) - f(Z^{(i)}))^2 | Z]$ is the variance proxy

Efron-Stein inequality bounds the deviation of a function in terms of local variance estimates obtained by changing one variable at a time

Paulin, Mackey and Tropp [Paulin et al., 2016] generalized Efron-Stein inequalities for matrices

- Uses the method of exchangeable pairs [Stein, 1972, Chatterjee, 2005]

Matrix Efron-Stein inequality

For independent random variables $Z_1, \ldots, Z_n$, let $Z^{(i)}$ denote $Z_1, \ldots, Z_{i-1}, \widetilde{Z}_i, Z_{i+1}, \ldots, Z_n$, where Z_i has been resampled to $\widetilde{Z}_i$

Matrix Efron-Stein inequality [Paulin et al., 2016]

Let $H(Z)$ be a Hermitian matrix valued function of independent random variables $Z = (Z_1, \ldots, Z_n)$ with $\mathbb{E}[\|H\|] < \infty$. Then, for each natural number $t \geq 1$,

$$\mathbb{E} \operatorname{tr}\left[(H - \mathbb{E} H)^{2t}\right] \;\leq\; (4t-2)^t \cdot \mathbb{E} \operatorname{tr}\left[V^t\right]$$

where $V(Z)$ is the variance proxy defined as

$$V(Z) := \frac{1}{2} \cdot \sum_{i=1}^{n} \mathbb{E}[(H(Z) - H(Z^{(i)}))^2 \mid Z]$$

Our approach

For a polynomial random matrix H, we interpret the variance proxy V in the matrix Efron-Stein inequality as obtained from partial derivatives of H

Our approach

For a polynomial random matrix H, we interpret the variance proxy V in the matrix Efron-Stein inequality as obtained from partial derivatives of H

With this alternate view,

- If we start with a polynomial random matrix, one application of matrix Efron-Stein reduces the degree by 1

Our approach

For a polynomial random matrix H, we interpret the variance proxy V in the matrix Efron-Stein inequality as obtained from partial derivatives of H

With this alternate view,

- If we start with a polynomial random matrix, one application of matrix Efron-Stein reduces the degree by 1
- We recursively apply it until we end up with deterministic matrices, which we can directly study

Our approach

For a polynomial random matrix H, we interpret the variance proxy V in the matrix Efron-Stein inequality as obtained from partial derivatives of H

With this alternate view,

- If we start with a polynomial random matrix, one application of matrix Efron-Stein reduces the degree by 1
- We recursively apply it until we end up with deterministic matrices, which we can directly study

We present a general framework based on this idea
This framework recovers known bounds in literature (up to lower order terms)

Our results

Theorem: Rademacher recursion

Let $F : \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ be a matrix valued polynomial function of degree at most d. Then, for each natural number $t \geq 1$,

$$\mathbb{E}\,\|F - \mathbb{E}\,F\|_{2t}^{2t} \;\leq \sum_{1\leq a+b\leq d} (16td)^{(a+b)\cdot t} \cdot \|\mathbb{E}\,F_{a,b}\|_{2t}^{2t}$$

where $F_{a,b}$ is a matrix of partial derivatives indexed by the sets $\mathcal{I} \times \binom{[n]}{a}$ and $\mathcal{J} \times \binom{[n]}{b}$ with

$$F_{a,b}[(\cdot,\alpha),(\cdot,\beta)] = \begin{cases} \nabla_{\alpha+\beta}(F) & \text{if } \alpha\cdot\beta = 0 \\ 0 & \text{otherwise} \end{cases}$$

Main takeaway: Reduces random matrix concentration to studying deterministic matrices.

Visualizing $F_{0,1}$

Let's visualize $F_{0,1}$ for clarity. Suppose

$$F = \begin{array}{rcl} & \text{column } J & \\ & \downarrow & \\ \text{row } I \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & F_{I,J}(Z) & \cdots \\ & \vdots & \end{pmatrix} & r \text{ rows} \\ & c \text{ columns} & \end{array}$$

Visualizing $F_{0,1}$

Let's visualize $F_{0,1}$ for clarity. Suppose

$$F = \begin{array}{ccc} & \text{column } J & \\ & \downarrow & \\ \text{row } I \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & F_{I,J}(Z) & \cdots \\ & \vdots & \end{pmatrix} & r \text{ rows} \\ & c \text{ columns} & \end{array}$$

Then,

$$F_{0,1} = \begin{array}{ccc} & \text{column } (J, \{i\}) & \\ & \downarrow & \\ \text{row } I \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & \nabla_{\mathbf{e}_i} F_{I,J}(Z) & \cdots \\ & \vdots & \end{pmatrix} & r \text{ rows} \\ & cn \text{ columns} & \end{array}$$

Rademacher recursion - Proof sketch

Let $F : \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ be an Hermitian matrix valued polynomial function

Assume $\mathbb{E}\, F = 0$, matrix Efron-Stein: $\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] \leq (4t-2)^t\, \mathbb{E}\,\mathrm{tr}\left[V^t\right]$

Rademacher recursion - Proof sketch

Let $F : \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ be an Hermitian matrix valued polynomial function

Assume $\mathbb{E}\, F = 0$, matrix Efron-Stein: $\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] \le (4t-2)^t\, \mathbb{E}\,\mathrm{tr}\left[V^t\right]$

$$V = \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[(F(Z) - F(Z^{(i)}))^2 \mid Z]$$

Rademacher recursion - Proof sketch

Let $F : \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ be an Hermitian matrix valued polynomial function

Assume $\mathbb{E}\, F = 0$, matrix Efron-Stein: $\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] \leq (4t-2)^t\,\mathbb{E}\,\mathrm{tr}\left[V^t\right]$

$$\begin{aligned} V &= \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}[(F(Z)-F(Z^{(i)}))^2 \mid Z] \\ &= \frac{1}{2}\sum_{i=1}^{n}\mathbb{E}[((Z_i-\widetilde{Z}_i) \;\cdot\; \underbrace{\nabla_{\mathbf{e}_i}F(Z)}_{\text{no randomness}}\;)^2 \mid Z] \end{aligned}$$

Rademacher recursion - Proof sketch

Let $F : \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ be an Hermitian matrix valued polynomial function
Assume $\mathbb{E}\, F = 0$, matrix Efron-Stein: $\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] \le (4t-2)^t\, \mathbb{E}\,\mathrm{tr}\left[V^t\right]$

$$\begin{aligned} V &= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[(F(Z) - F(Z^{(i)}))^2 \mid Z] \\ &= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[((Z_i - \widetilde{Z}_i) \quad \cdot \quad \underbrace{\nabla_{\mathbf{e}_i} F(Z)}_{\text{no randomness}} \quad)^2 \mid Z] \\ &= \frac{1}{2}\sum_{i=1}^{n} \underbrace{\mathbb{E}[(Z_i - \widetilde{Z}_i)^2 | Z]}_{=2} \cdot (\nabla_{\mathbf{e}_i} F(Z))^2 \end{aligned}$$

Rademacher recursion - Proof sketch

Let $F : \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ be an Hermitian matrix valued polynomial function

Assume $\mathbb{E}\, F = 0$, matrix Efron-Stein: $\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] \le (4t-2)^t\, \mathbb{E}\,\mathrm{tr}\left[V^t\right]$

$$
\begin{aligned}
V &= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[(F(Z) - F(Z^{(i)}))^2 \mid Z] \\
&= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[((Z_i - \widetilde{Z}_i) \quad \cdot \quad \underbrace{\nabla_{\mathbf{e}_i} F(Z)}_{\text{no randomness}} \quad)^2 \mid Z] \\
&= \frac{1}{2}\sum_{i=1}^{n} \underbrace{\mathbb{E}[(Z_i - \widetilde{Z}_i)^2 | Z]}_{=2} \cdot (\nabla_{\mathbf{e}_i} F(Z))^2 \\
&= \sum_{i=1}^{n} (\nabla_{\mathbf{e}_i} F(Z))^2
\end{aligned}
$$

Rademacher recursion - Proof sketch

Let $F : \{-1, 1\}^n \to \mathbb{R}^{\mathcal{I} \times \mathcal{J}}$ be an Hermitian matrix valued polynomial function

Assume $\mathbb{E}\, F = 0$, matrix Efron-Stein: $\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] \le (4t-2)^t\, \mathbb{E}\,\mathrm{tr}\left[V^t\right]$

$$\begin{aligned}
V &= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[(F(Z) - F(Z^{(i)}))^2 \mid Z] \\
&= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[((Z_i - \widetilde{Z}_i) \;\cdot\; \underbrace{\nabla_{\mathbf{e}_i} F(Z)}_{\text{no randomness}}\;)^2 \mid Z] \\
&= \frac{1}{2}\sum_{i=1}^{n} \underbrace{\mathbb{E}[(Z_i - \widetilde{Z}_i)^2 | Z]}_{=2} \cdot (\nabla_{\mathbf{e}_i} F(Z))^2 \\
&= \sum_{i=1}^{n} (\nabla_{\mathbf{e}_i} F(Z))^2 \\
&= F_{0,1} F_{0,1}^{\mathsf{T}}
\end{aligned}$$

Rademacher recursion - Proof sketch

So far

$$\begin{aligned}\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] &\leq (4t-2)^t\,\mathbb{E}\,\mathrm{tr}\left[V^t\right]\\ &= O(t)^t\,\mathbb{E}\,\|F_{0,1}\|_{2t}^{2t}\end{aligned}$$

Rademacher recursion - Proof sketch

So far

$$\begin{aligned}\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] &\leq (4t-2)^t\,\mathbb{E}\,\mathrm{tr}\left[V^t\right]\\ &= O(t)^t\,\mathbb{E}\,\|F_{0,1}\|_{2t}^{2t}\\ &= O(t)^t(\|\underbrace{\mathbb{E}\,F_{0,1}}_{\textit{deterministic}}\|_{2t}^{2t}+\underbrace{\mathbb{E}\,\|F_{0,1}-\mathbb{E}\,F_{0,1}\|_{2t}^{2t}}_{\textit{recurse}!})\end{aligned}$$

Rademacher recursion - Proof sketch

So far

$$\begin{aligned} \mathbb{E}\operatorname{tr}\left[F^{2t}\right] &\leq (4t-2)^t\,\mathbb{E}\operatorname{tr}\left[V^t\right] \\ &= O(t)^t\,\mathbb{E}\,\|F_{0,1}\|_{2t}^{2t} \\ &= O(t)^t(\|\underbrace{\mathbb{E}\,F_{0,1}}_{deterministic}\|_{2t}^{2t} + \underbrace{\mathbb{E}\,\|F_{0,1} - \mathbb{E}\,F_{0,1}\|_{2t}^{2t}}_{recurse!}) \end{aligned}$$

For non-Hermitian F, we instead use Hermitian dilation $\begin{bmatrix} 0 & F \\ F^\intercal & 0 \end{bmatrix}$

So, along with $F_{0,1}$, we get $F_{1,0}$ as well.

Rademacher recursion - Proof sketch

So far

$$\begin{aligned}\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] &\leq (4t-2)^t\,\mathbb{E}\,\mathrm{tr}\left[V^t\right]\\ &= O(t)^t\,\mathbb{E}\,\|F_{0,1}\|_{2t}^{2t}\\ &= O(t)^t(\|\underbrace{\mathbb{E}\,F_{0,1}}_{deterministic}\|_{2t}^{2t} + \underbrace{\mathbb{E}\,\|F_{0,1} - \mathbb{E}\,F_{0,1}\|_{2t}^{2t}}_{recurse!})\end{aligned}$$

For non-Hermitian F, we instead use Hermitian dilation $\begin{bmatrix} 0 & F \\ F^\intercal & 0 \end{bmatrix}$

So, along with $F_{0,1}$, we get $F_{1,0}$ as well.

$$\begin{aligned}\mathbb{E}\,\mathrm{tr}\left[F^{2t}\right] \leq O(t)^t(&\|\mathbb{E}\,F_{0,1}\|_{2t}^{2t} + \|\mathbb{E}\,F_{1,0}\|_{2t}^{2t}\\ &+ \mathbb{E}\,\|F_{0,2}\|_{2t}^{2t} + \mathbb{E}\,\|F_{1,1}\|_{2t}^{2t} + \mathbb{E}\,\|F_{2,0}\|_{2t}^{2t})\end{aligned}$$

Rademacher recursion - Proof sketch

So far

$$\begin{aligned}\mathbb{E}\operatorname{tr}\left[F^{2t}\right] &\leq (4t-2)^t\,\mathbb{E}\operatorname{tr}\left[V^t\right]\\ &= O(t)^t\,\mathbb{E}\,\|F_{0,1}\|_{2t}^{2t}\\ &= O(t)^t(\|\underbrace{\mathbb{E}\,F_{0,1}}_{\textit{deterministic}}\|_{2t}^{2t} + \underbrace{\mathbb{E}\,\|F_{0,1}-\mathbb{E}\,F_{0,1}\|_{2t}^{2t}}_{\textit{recurse}!})\end{aligned}$$

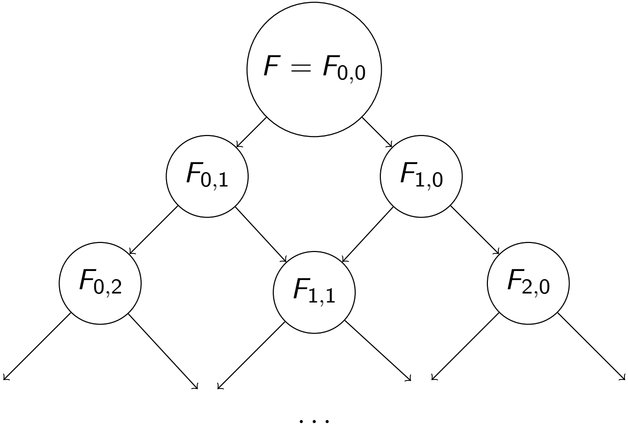
For non-Hermitian F, we instead use Hermitian dilation $\begin{bmatrix} 0 & F \\ F^\intercal & 0 \end{bmatrix}$

So, along with $F_{0,1}$, we get $F_{1,0}$ as well.

$$\begin{aligned}\mathbb{E}\operatorname{tr}\left[F^{2t}\right] &\leq O(t)^t(\|\mathbb{E}\,F_{0,1}\|_{2t}^{2t} + \|\mathbb{E}\,F_{1,0}\|_{2t}^{2t}\\ &\qquad + \mathbb{E}\,\|F_{0,2}\|_{2t}^{2t} + \mathbb{E}\,\|F_{1,1}\|_{2t}^{2t} + \mathbb{E}\,\|F_{2,0}\|_{2t}^{2t})\\ &\leq \ldots \text{ recurse until we get deterministic matrices}\end{aligned}$$

This is essentially our main result for Rademacher variables

Visualizing the recursion



In each layer, we extract out the expectation and apply matrix Efron-Stein on a new centered random matrix

Because the polynomial degree is bounded, this will stop

Our results

Theorem: Rademacher recursion

Let $F : \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ be a matrix valued polynomial function of degree at most d. Then, for each natural number $t \geq 1$,

$$\mathbb{E}\,\|F - \mathbb{E}\,F\|_{2t}^{2t} \;\leq\; \sum_{1\leq a+b\leq d} (16td)^{(a+b)\cdot t} \cdot \|\mathbb{E}\,F_{a,b}\|_{2t}^{2t}$$

where $F_{a,b}$ is a matrix of partial derivatives indexed by the sets $\mathcal{I} \times \binom{[n]}{a}$ and $\mathcal{J} \times \binom{[n]}{b}$ with

$$F_{a,b}[(\cdot,\alpha),(\cdot,\beta)] = \begin{cases} \nabla_{\alpha+\beta}(F) & \text{if } \alpha\cdot\beta = 0 \\ 0 & \text{otherwise} \end{cases}$$

Main takeaway: Reduces random matrix concentration to studying deterministic matrices.

Graph matrices

We show an application to obtain concentration for “graph matrices”

Graph matrices

We show an application to obtain concentration for “graph matrices”

Graph matrices: Special class of polynomial random matrices, that have useful properties and can be represented diagramatically

Graph matrices

We show an application to obtain concentration for “graph matrices”

Graph matrices: Special class of polynomial random matrices, that have useful properties and can be represented diagramatically

Matrices represented by “shapes” τ (which are just other smaller graphs)

Graph matrices

We show an application to obtain concentration for “graph matrices”

Graph matrices: Special class of polynomial random matrices, that have useful properties and can be represented diagramatically

Matrices represented by “shapes” τ (which are just other smaller graphs)

Useful to design spectral algorithms and to study high degree SoS lower bounds

- Studied by Ahn, Medarametla and Potechin [Medarametla and Potechin, 2016, Ahn et al., 2020]
- Closely related to tensor networks [Moitra and Wein, 2019]

Graph matrices

Fix an underlying graph $G \sim \mathcal{G}_{n,1/2}$

Graph matrices

Fix an underlying graph $G \sim \mathcal{G}_{n,1/2}$

A shape τ is another graph that encodes a matrix M_τ

Graph matrices

Fix an underlying graph $G \sim \mathcal{G}_{n,1/2}$

A shape τ is another graph that encodes a matrix M_τ

Shape τ has two special subsets of ordered vertices U_τ, V_τ, that encode row and column indices

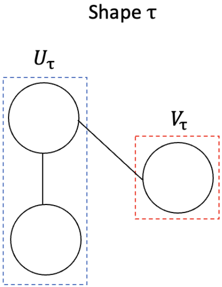
Graph matrices

Fix an underlying graph $G \sim \mathcal{G}_{n,1/2}$

A shape τ is another graph that encodes a matrix M_τ

Shape τ has two special subsets of ordered vertices U_τ, V_τ, that encode row and column indices

Example:



Graph matrices

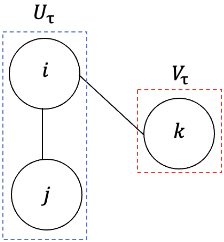
To define M_τ, we consider injective “realizations” of τ and fill in the entries of the matrix accordingly

A realization is an injective map from vertices of τ to $[n]$

Graph matrices

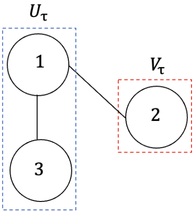
To define M_τ, we consider injective "realizations" of τ and fill in the entries of the matrix accordingly
A realization is an injective map from vertices of τ to $[n]$

For example, suppose i, j, k are distinct elements of $[n]$, then a realization mapping vertices to i, j, k is



Graph matrices

Realizations of U_τ, V_τ correspond to row and column indices



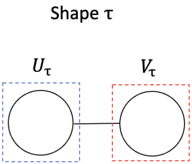
corresponds to row (1, 3) and column (2)

To define M_τ, we go over all possible realizations and assign entries accordingly

Edges correspond to input variables, so $G_{1,2}$ and $G_{1,3}$ in this case

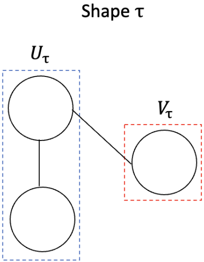
For ease of calculations, some works use a different definition where we go over all distinct realizations

Example graph matrices



corresponds to

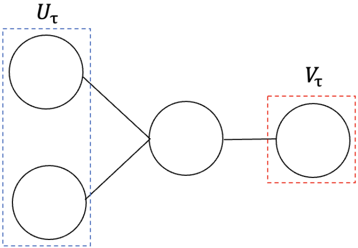
$$M_\tau = \begin{array}{rcl} & \text{column } (j) & \\ & \downarrow & \\ \text{row } (i) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & G_{ij} & \cdots \\ & \vdots & \end{pmatrix} & n \text{ rows} \\ & n \text{ columns} & \end{array}$$



corresponds to

$$M_\tau = \begin{array}{rcl} & \text{column } (k) & \\ & \downarrow & \\ \text{row } (i,j) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & G_{ij} G_{ik} & \cdots \\ & \vdots & \end{pmatrix} & O(n^2) \text{ rows} \\ & \text{n columns} & \end{array}$$

Example graph matrices



corresponds to

$$M_\tau = \begin{array}{cc} & \text{column } (k) \\ & \downarrow \\ \text{row } (i,j) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & \sum_{l \in [n]-\{i,j,k\}} G_{il} G_{jl} G_{kl} & \cdots \\ & \vdots & \end{pmatrix} O(n^2) \text{ rows} \\ & \text{n columns} \end{array}$$

Concentration of graph matrices

Obtaining norm bounds on graph matrices is an important task in SoS analysis and in our work

Concentration of graph matrices

Obtaining norm bounds on graph matrices is an important task in SoS analysis and in our work

Norm bounds have been obtained by Ahn, Medarametla and Potechin [Medarametla and Potechin, 2016, Ahn et al., 2020]

Concentration of graph matrices

Obtaining norm bounds on graph matrices is an important task in SoS analysis and in our work

Norm bounds have been obtained by Ahn, Medarametla and Potechin [Medarametla and Potechin, 2016, Ahn et al., 2020]

In particular, they use the trace method

Concentration of graph matrices

Obtaining norm bounds on graph matrices is an important task in SoS analysis and in our work

Norm bounds have been obtained by Ahn, Medarametla and Potechin [Medarametla and Potechin, 2016, Ahn et al., 2020]

In particular, they use the trace method

Nontrivial arguments show that a surprising combinatorial structure called the minimum vertex separator of the shape τ dictates the norm bound of the graph matrix M_τ

Concentration of graph matrices

Obtaining norm bounds on graph matrices is an important task in SoS analysis and in our work

Norm bounds have been obtained by Ahn, Medarametla and Potechin [Medarametla and Potechin, 2016, Ahn et al., 2020]

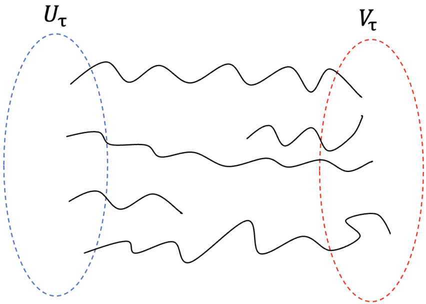
In particular, they use the trace method

Nontrivial arguments show that a surprising combinatorial structure called the minimum vertex separator of the shape τ dictates the norm bound of the graph matrix M_τ

We obtain this combinatorial structure alternatively through our framework

Minimum vertex separator

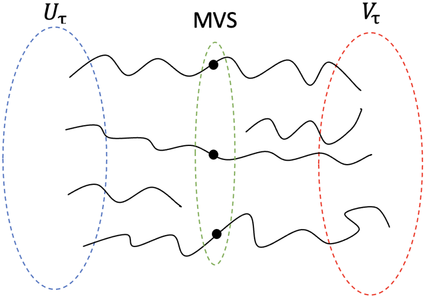
Picture a shape τ



Minimum vertex separator (MVS): Minimum set of vertices whose removal disconnects U_τ from V_τ

Minimum vertex separator

Pictorial representation of MVS



Norm bound on graph matrices

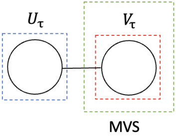
Theorem [Medarametla and Potechin, 2016]

For a shape τ with no degree-0 vertices outside $U_\tau \cup V_\tau$, $\|M_\tau\| \leq \tilde{O}(\sqrt{n}^{|V(\tau)|-|S|})$ w.h.p. where S is an MVS

Examples:

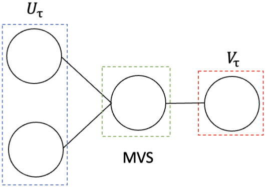
$$\begin{array}{ccc} & \text{column } (j) & \\ & \downarrow & \\ \text{row } (i) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots & G_{ij} & \cdots \\ & \vdots & \end{pmatrix} & n \text{ rows} \\ & n \text{ columns} & \end{array}$$

Norm $\tilde{O}(\sqrt{n})$



$$\begin{array}{ccc} & \text{column } (k) & \\ & \downarrow & \\ \text{row } (i,j) \rightarrow & \begin{pmatrix} & \vdots & \\ \cdots \sum_{l \in [n]-\{i,j,k\}} G_{il} G_{jl} G_{kl} \cdots \\ & \vdots & \end{pmatrix} & O(n^2) \text{ rows} \\ & \text{n columns} & \end{array}$$

Norm $\tilde{O}(n^{1.5})$



Norm bound on graph matrices

Theorem [Medarametla and Potechin, 2016]

For a shape τ with no degree-0 vertices outside $U_\tau \cup V_\tau$, $\|M_\tau\| \leq \tilde{O}(\sqrt{n}^{|V(\tau)|-|S|})$ w.h.p. where S is an MVS

Norm bound on graph matrices

Theorem [Medarametla and Potechin, 2016]

For a shape τ with no degree-0 vertices outside $U_\tau \cup V_\tau$,
$\|M_\tau\| \leq \tilde{O}(\sqrt{n}^{|V(\tau)|-|S|})$ w.h.p. where S is an MVS

Proof idea

Their proof applies the trace method and makes some beautiful observations based on Menger's theorem to obtain the norm bound

Norm bound on graph matrices

Theorem [Medarametla and Potechin, 2016]

For a shape τ with no degree-0 vertices outside $U_\tau \cup V_\tau$, $\|M_\tau\| \leq \tilde{O}(\sqrt{n}^{|V(\tau)|-|S|})$ w.h.p. where S is an MVS

Proof idea

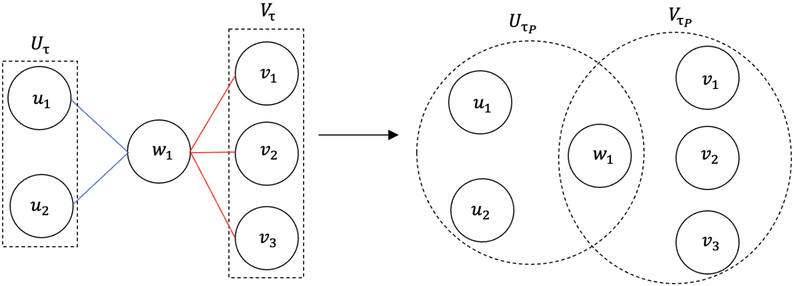
Their proof applies the trace method and makes some beautiful observations based on Menger's theorem to obtain the norm bound

We instead apply our framework

- Fix τ, then $F = M_\tau$ is our input random polynomial matrix
- We just need to think about $\mathbb{E}\, F_{a,b}$ for $a+b$ being the degree of F
- By appropriately renaming the rows and columns, we can think of them as graph matrices as well!

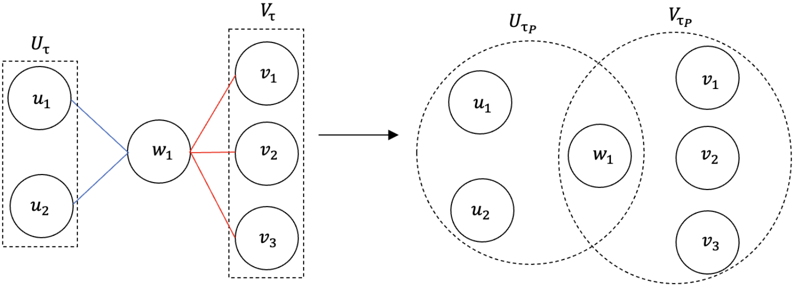
An example of a final term

One possible term $\mathbb{E}\, F_{2,3}$ in our inequality



An example of a final term

One possible term $\mathbb{E}\, F_{2,3}$ in our inequality



Each term $\mathbb{E}\, F_{a,b}$ can be viewed as

- Pick a edges, delete them and move their incident vertices to U_τ
- Do the same for the remaining b edges but with V_τ instead
- Obtain a deterministic matrix M_{τ_P}.
- Easy to bound their norm, just $\sqrt{n}^{|V(\tau)|-|U_{\tau_P}\cap V_{\tau_P}|}$
- They are governed by number of “common vertices in U and V”

Norm bounds on graph matrices

We just do this for all a, b and apply our inequality.

Norm bounds on graph matrices

We just do this for all a, b and apply our inequality.

End up with a norm bound of the form $\widetilde{O}(\sqrt{n}^{|V(\tau)|-|S|})$ where S is the number of "common vertices in U and V"

Norm bounds on graph matrices

We just do this for all a, b and apply our inequality.

End up with a norm bound of the form $\widetilde{O}(\sqrt{n}^{|V(\tau)|-|S|})$ where S is the number of "common vertices in U and V"

Observation

S obtained this way must be a vertex separator of the original τ

Norm bounds on graph matrices

We just do this for all a, b and apply our inequality.

End up with a norm bound of the form $\widetilde{O}(\sqrt{n}^{|V(\tau)|-|S|})$ where S is the number of "common vertices in U and V"

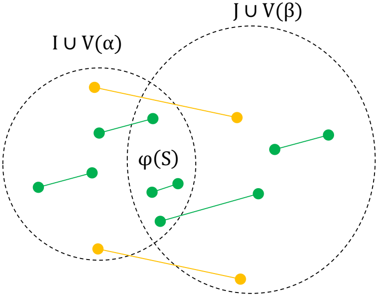
Observation

S obtained this way must be a vertex separator of the original τ

If we prove this, then we are done since we get a final norm bound of $\sqrt{n}^{|V(\tau)|-|S|}$ where S is a MVS, just like prior works derived.

Proof of why S is a vertex separator

Proof by picture:



Green edges can occur in τ but orange edges cannot
Therefore, we are done

The case of non-Rademacher variabels

Our ideas worked well for Rademacher random variables Z_i.

The case of non-Rademacher variabels

Our ideas worked well for Rademacher random variables Z_i.

But in many applications, we have non-Rademacher Z_i, such as graph matrices when $G \sim \mathcal{G}_{n,p}$

- These are useful to study algorithms on sparse Erdős-Rényi graphs.

The case of non-Rademacher variabels

Our ideas worked well for Rademacher random variables Z_i.

But in many applications, we have non-Rademacher Z_i, such as graph matrices when $G \sim \mathcal{G}_{n,p}$

- These are useful to study algorithms on sparse Erdős-Rényi graphs.

We could attempt the same recursion idea but

$$\mathbb{E}[(Z_i - \widetilde{Z}_i)^2 | Z] = 1 + Z_i^2 \neq 2$$

The polynomial degree doesn't decrease and the recursion stalls!

A generalization for non-Rademacher variabels

To get around this, we present a generalized version of our theorem for independent variables

- In the Rademacher case, random matrix concentration was reduced to studying deterministic matrices
- In the general case, it reduces to scalar concentration
- Main technical contribution of our work

A generalization for non-Rademacher variabels

To get around this, we present a generalized version of our theorem for independent variables

- In the Rademacher case, random matrix concentration was reduced to studying deterministic matrices
- In the general case, it reduces to scalar concentration
- Main technical contribution of our work

Proof idea

- Extend the proof of the matrix Efron-Stein inequality [Paulin et al., 2016]
- Use the method of exchangeable pairs but with explicit polynomial kernels that we construct

A generalization for non-Rademacher variabels

To get around this, we present a generalized version of our theorem for independent variables

- In the Rademacher case, random matrix concentration was reduced to studying deterministic matrices
- In the general case, it reduces to scalar concentration
- Main technical contribution of our work

Proof idea

- Extend the proof of the matrix Efron-Stein inequality [Paulin et al., 2016]
- Use the method of exchangeable pairs but with explicit polynomial kernels that we construct

We apply them to obtain norm bounds for graph matrices in the sparse setting

Table of Contents

The Sum of Squares hierarchy

SoS is a proof system for reasoning about polynomial inequalities

The Sum of Squares hierarchy

SoS is a proof system for reasoning about polynomial inequalities

Had been studied by Hilbert and Minkowski back in 1900s

The Sum of Squares hierarchy

SoS is a proof system for reasoning about polynomial inequalities

Had been studied by Hilbert and Minkowski back in 1900s

[Parrilo, 2000] and [Lasserre, 2001] independently showed algorithmic implications of SoS proofs

The Sum of Squares hierarchy

SoS is a proof system for reasoning about polynomial inequalities

Had been studied by Hilbert and Minkowski back in 1900s

[Parrilo, 2000] and [Lasserre, 2001] independently showed algorithmic implications of SoS proofs

The SoS hierarchy is a family of semidefinite programs parameterized by the degree or level d

The Sum of Squares hierarchy

SoS is a proof system for reasoning about polynomial inequalities

Had been studied by Hilbert and Minkowski back in 1900s

[Parrilo, 2000] and [Lasserre, 2001] independently showed algorithmic implications of SoS proofs

The SoS hierarchy is a family of semidefinite programs parameterized by the degree or level d

The programs get stronger as d increases
Degree-d SoS can be solved in $n^{O(d)}$ time (ignoring bit complexity issues), where n is the input size

The Sum of Squares hierarchy

SoS is a proof system for reasoning about polynomial inequalities

Had been studied by Hilbert and Minkowski back in 1900s

[Parrilo, 2000] and [Lasserre, 2001] independently showed algorithmic implications of SoS proofs

The SoS hierarchy is a family of semidefinite programs parameterized by the degree or level d

The programs get stronger as d increases
Degree-d SoS can be solved in $n^{O(d)}$ time (ignoring bit complexity issues), where n is the input size

Constant degree d corresponds to polynomial runtime
In this work, our lower bounds focus on $d \approx n^{\epsilon}$, corresponding to subexponential runtime

The pseudoexpectation operator

Consider a polynomial system $p_1(x) = 0, \ldots, p_m(x) = 0$ on n variables $x_1, \ldots, x_n$, where $deg(p_i) \leq d$ for all i

The pseudoexpectation operator

Consider a polynomial system $p_1(x) = 0, \ldots, p_m(x) = 0$ on n variables $x_1, \ldots, x_n$, where $deg(p_i) \leq d$ for all i

Pseudoexpectation operator

A degree-d pseudoexpectation operator $\widetilde{\mathbb{E}} : \mathbb{R}^{\leq d}[x] \to \mathbb{R}$ is a linear operator such that

- $\widetilde{\mathbb{E}}[1] = 1$
- $\widetilde{\mathbb{E}}[p_i f] = 0$ for all $f \in \mathbb{R}^{\leq d}[x]$ such that $deg(p_i f) \leq d$
- $\widetilde{\mathbb{E}}[f^2] \geq 0$ for all $f \in \mathbb{R}^{\leq d}[x]$ such that $deg(f^2) \leq d$

The SoS hierarchy

Consider a program where we want to maximize $f(x)$ subject to $p_1(x) = 0, \ldots, p_m(x) = 0$

The SoS hierarchy

Consider a program where we want to maximize $f(x)$ subject to $p_1(x) = 0, \ldots, p_m(x) = 0$

Degree-d SoS

Maximize $\widetilde{\mathbb{E}}[f]$ over all valid pseudoexpectation operators $\widetilde{\mathbb{E}}$ satisfying the polynomial system

The SoS hierarchy

Consider a program where we want to maximize $f(x)$ subject to $p_1(x) = 0, \ldots, p_m(x) = 0$

Degree-d SoS

Maximize $\widetilde{\mathbb{E}}[f]$ over all valid pseudoexpectation operators $\widetilde{\mathbb{E}}$ satisfying the polynomial system

Can use semidefinite programming to do this in time $(m+n)^{O(d)}$

The SoS hierarchy

Consider a program where we want to maximize $f(x)$ subject to $p_1(x) = 0, \ldots, p_m(x) = 0$

Degree-d SoS

Maximize $\widetilde{\mathbb{E}}[f]$ over all valid pseudoexpectation operators $\widetilde{\mathbb{E}}$ satisfying the polynomial system

Can use semidefinite programming to do this in time $(m+n)^{O(d)}$

Suppose D is a distribution over optimal solutions x^*
Then $\widetilde{\mathbb{E}}$ defined as $\widetilde{\mathbb{E}}[g(x)] = \mathbb{E}_{x^* \sim D}[g(x^*)]$ is a valid pseudoexpectation operator

The SoS hierarchy

Consider a program where we want to maximize $f(x)$ subject to $p_1(x) = 0, \ldots, p_m(x) = 0$

Degree-d SoS

Maximize $\widetilde{\mathbb{E}}[f]$ over all valid pseudoexpectation operators $\widetilde{\mathbb{E}}$ satisfying the polynomial system

Can use semidefinite programming to do this in time $(m+n)^{O(d)}$

Suppose D is a distribution over optimal solutions x^*
Then $\widetilde{\mathbb{E}}$ defined as $\widetilde{\mathbb{E}}[g(x)] = \mathbb{E}_{x^* \sim D}[g(x^*)]$ is a valid pseudoexpectation operator

Hence, degree-d SoS is a relaxation of the polynomial program and can be used for certifying upper bounds on the optimum

Example - Maximum clique

Given a graph G, certify an upper bound on the size of the maximum clique.

Example - Maximum clique

Given a graph G, certify an upper bound on the size of the maximum clique.

We consider the program

$$\begin{aligned} \text{Maximize } \sum_{i=1}^{n} x_i & \\ x_i x_j = 0 \qquad & \forall (i,j) \notin E(G) \\ x_i^2 = x_i \qquad & \forall i \in [n] \end{aligned}$$

x_i indicates whether vertex i is in the clique

Example - Maximum clique

Given a graph G, certify an upper bound on the size of the maximum clique.

We consider the program

$$\begin{aligned} \text{Maximize } \sum_{i=1}^{n} x_i & \\ x_i x_j = 0 \qquad & \forall (i,j) \notin E(G) \\ x_i^2 = x_i \qquad & \forall i \in [n] \end{aligned}$$

x_i indicates whether vertex i is in the clique

Known that the degree-2 SoS relaxation of the above program certifies an upper bound of $O(\sqrt{n})$ w.h.p. when $G \sim \mathcal{G}_{n,1/2}$

Example - Maximum clique

Given a graph G, certify an upper bound on the size of the maximum clique.

We consider the program

$$\begin{aligned} \text{Maximize } \sum_{i=1}^{n} x_i & \\ x_i x_j = 0 & \qquad \forall (i,j) \notin E(G) \\ x_i^2 = x_i & \qquad \forall i \in [n] \end{aligned}$$

x_i indicates whether vertex i is in the clique

Known that the degree-2 SoS relaxation of the above program certifies an upper bound of $O(\sqrt{n})$ w.h.p. when $G \sim \mathcal{G}_{n,1/2}$

Can higher degree SoS do better? Such questions are our main focus

SoS lower bounds

Why?

- SoS achieves state-of-the-art algorithmic guarantees for many problems in in optimization
  - E.g., MaxCut [Goemans and Williamson, 1995], Sparsest cut [Arora et al., 2004], Tensor PCA [Hopkins et al., 2015]

SoS lower bounds

Why?

- SoS achieves state-of-the-art algorithmic guarantees for many problems in in optimization
  - E.g., MaxCut [Goemans and Williamson, 1995], Sparsest cut [Arora et al., 2004], Tensor PCA [Hopkins et al., 2015]
- Strong evidence of hardness of computation

SoS lower bounds

Why?

- SoS achieves state-of-the-art algorithmic guarantees for many problems in in optimization
    - E.g., MaxCut [Goemans and Williamson, 1995], Sparsest cut [Arora et al., 2004], Tensor PCA [Hopkins et al., 2015]
- Strong evidence of hardness of computation

Some prior works

- SoS lower bounds for k-XOR, k-SAT [Grigoriev, 2001] which were reduced to other problems [Tulsiani, 2009]

SoS lower bounds

Why?

- SoS achieves state-of-the-art algorithmic guarantees for many problems in in optimization
  - E.g., MaxCut [Goemans and Williamson, 1995], Sparsest cut [Arora et al., 2004], Tensor PCA [Hopkins et al., 2015]
- Strong evidence of hardness of computation

Some prior works

- SoS lower bounds for k-XOR, k-SAT [Grigoriev, 2001] which were reduced to other problems [Tulsiani, 2009]
- Max *k*-CSPs [Kothari, Mori, O'Donnell, Witmer, 2017]

SoS lower bounds

Why?

- SoS achieves state-of-the-art algorithmic guarantees for many problems in in optimization
  - E.g., MaxCut [Goemans and Williamson, 1995], Sparsest cut [Arora et al., 2004], Tensor PCA [Hopkins et al., 2015]
- Strong evidence of hardness of computation

Some prior works

- SoS lower bounds for k-XOR, k-SAT [Grigoriev, 2001] which were reduced to other problems [Tulsiani, 2009]
- Max k-CSPs [Kothari, Mori, O'Donnell, Witmer, 2017]
- Maximum clique on random graphs [Barak, Hopkins, Kelner, Kothari, Moitra, Potechin, 2016]

SoS lower bounds

Why?

- SoS achieves state-of-the-art algorithmic guarantees for many problems in in optimization
  - E.g., MaxCut [Goemans and Williamson, 1995], Sparsest cut [Arora et al., 2004], Tensor PCA [Hopkins et al., 2015]
- Strong evidence of hardness of computation

Some prior works

- SoS lower bounds for k-XOR, k-SAT [Grigoriev, 2001] which were reduced to other problems [Tulsiani, 2009]
- Max *k*-CSPs [Kothari, Mori, O'Donnell, Witmer, 2017]
- Maximum clique on random graphs [Barak, Hopkins, Kelner, Kothari, Moitra, Potechin, 2016]
- These are highly nontrivial results building on years of research and use a lot of interesting ideas

SoS lower bounds in this work

In our lower bounds, we broadly build on the techniques in [Barak et al., 2016]

SoS lower bounds in this work

In our lower bounds, we broadly build on the techniques in [Barak et al., 2016]

But the problems we study pose their own set of challenges, e.g.,

- Sherrington-Kirkpatrick requires formulating and studying a dual problem, and brings additional conceptual challenges
- Sparse PCA deals with Gaussian inputs instead of boolean inputs considered in prior works

SoS lower bounds in this work

In our lower bounds, we broadly build on the techniques in [Barak et al., 2016]

But the problems we study pose their own set of challenges, e.g.,

- Sherrington-Kirkpatrick requires formulating and studying a dual problem, and brings additional conceptual challenges
- Sparse PCA deals with Gaussian inputs instead of boolean inputs considered in prior works

We will describe the Sherrington-Kirkpatrick lower bound next

An optimization task

Given a $n \times n$ matrix W, compute

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^{\intercal} W x$$

An optimization task

Given a $n \times n$ matrix W, compute

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^{\mathsf{T}} W x$$

Arises in Computer Science and Statistical Physics
Indeed, consider W distributed as

- Laplacian of a random graph: Maximum cut problem on random graphs
- $GOE(n)$: Random Hamiltonian of the celebrated Sherrington-Kirkpatrick model

Gaussian Orthogonal Ensemble $GOE(n)$

$GOE(n)$ is the distribution of $W = \frac{1}{\sqrt{2}}(A + A^T)$ where $A \in \mathbb{R}^{n \times n}$ with i.i.d. standard Gaussian entries

The case of $GOE(n)$ - The SK problem

Compute

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^\intercal W x$$

when $W \sim GOE(n)$

The case of $GOE(n)$ - The SK problem

Compute

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^\intercal W x$$

when $W \sim GOE(n)$

- The objective is the random Hamiltonian of the Sherrington-Kirkpatrick model from Statistical Physics
- Optimal value corresponds to the minimum energy of the system, up to sign and up to scaling

The case of $GOE(n)$ - The SK problem

Compute

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^\intercal W x$$

when $W \sim GOE(n)$

- The objective is the random Hamiltonian of the Sherrington-Kirkpatrick model from Statistical Physics
- Optimal value corresponds to the minimum energy of the system, up to sign and up to scaling
- This problem has deep connections to maximum cut on random graphs

The SK optimization problem

The true optimum and the Parisi constant

$$\lim_{n\to\infty} \mathop{\mathbb{E}}_{W\sim GOE(n)}[\frac{1}{n^{3/2}} OPT(W)] = 2P^*$$

The SK optimization problem

The true optimum and the Parisi constant

$$\lim_{n\to\infty} \mathop{\mathbb{E}}_{W\sim GOE(n)}[\frac{1}{n^{3/2}}OPT(W)] = 2P^*$$

Non-rigorous predictions of [Parisi, 1979, Crisanti and Rizzo, 2002], eventually formalized by [Talagrand, 2006, Panchenko, 2014] show

$$P^* \approx 0.7632$$

The SK optimization problem

The true optimum and the Parisi constant

$$\lim_{n\to\infty} \mathop{\mathbb{E}}_{W\sim GOE(n)}[\frac{1}{n^{3/2}}OPT(W)] = 2P^*$$

Non-rigorous predictions of [Parisi, 1979, Crisanti and Rizzo, 2002], eventually formalized by [Talagrand, 2006, Panchenko, 2014] show

$$P^* \approx 0.7632$$

[Montanari, 2019] essentially settled the optimization variant: For any $\epsilon > 0$, there is a polynomial time search algorithm that w.h.p. achieves a value of $(2P^* - \epsilon)n^{3/2}$, assuming a widely believed conjecture

The SK optimization problem

The true optimum and the Parisi constant

$$\lim_{n\to\infty} \mathop{\mathbb{E}}_{W\sim GOE(n)}[\frac{1}{n^{3/2}} OPT(W)] = 2P^*$$

Non-rigorous predictions of [Parisi, 1979, Crisanti and Rizzo, 2002], eventually formalized by [Talagrand, 2006, Panchenko, 2014] show

$$P^* \approx 0.7632$$

[Montanari, 2019] essentially settled the optimization variant: For any $\epsilon > 0$, there is a polynomial time search algorithm that w.h.p. achieves a value of $(2P^* - \epsilon)n^{3/2}$, assuming a widely believed conjecture

Let's move on to certification

The SK certification problem

Certify upper bounds on

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^\intercal W x$$

when $W \sim GOE(n)$.

The SK certification problem

Certify upper bounds on

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^\intercal W x$$

when $W \sim GOE(n)$.

In other words, design an efficient algorithm $\mathcal{A}$ that on input W outputs a certifiable $\mathcal{A}(W)$ such that

- $OPT(W) \leq \mathcal{A}(W)$
- On most instances, $\mathcal{A}(W)$ is *reasonably* close to $OPT(W)$

The spectral certificate

Consider the spectral algorithm $\mathcal{A}$ that outputs $\mathcal{A}(W) = \lambda_{max}(W)n$. Indeed,

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^\intercal W x \leq \lambda_{max}(W)n = \mathcal{A}(W)$$

The spectral certificate

Consider the spectral algorithm $\mathcal{A}$ that outputs $\mathcal{A}(W) = \lambda_{max}(W)n$.
Indeed,

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^\intercal W x \leq \lambda_{max}(W)n = \mathcal{A}(W)$$

How good is this?
From random matrix theory, w.h.p.,

$$\lambda_{max}(W) = (2 + o_n(1)) \cdot \sqrt{n}$$

So, this algorithm certifies w.h.p., $OPT(W) \leq (2 + o_n(1)) \cdot n^{3/2}$

The spectral certificate

Consider the spectral algorithm $\mathcal{A}$ that outputs $\mathcal{A}(W) = \lambda_{max}(W)n$.
Indeed,

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^\intercal W x \leq \lambda_{max}(W)n = \mathcal{A}(W)$$

How good is this?
From random matrix theory, w.h.p.,

$$\lambda_{max}(W) = (2 + o_n(1)) \cdot \sqrt{n}$$

So, this algorithm certifies w.h.p., $OPT(W) \leq (2 + o_n(1)) \cdot n^{3/2}$

Recall that the true optimum is $\approx 1.526 \cdot n^{3/2}$ w.h.p.

The spectral certificate

Consider the spectral algorithm $\mathcal{A}$ that outputs $\mathcal{A}(W) = \lambda_{max}(W)n$.
Indeed,

$$OPT(W) = \max_{x \in \{\pm 1\}^n} x^\intercal W x \leq \lambda_{max}(W)n = \mathcal{A}(W)$$

How good is this?
From random matrix theory, w.h.p.,

$$\lambda_{max}(W) = (2 + o_n(1)) \cdot \sqrt{n}$$

So, this algorithm certifies w.h.p., $OPT(W) \leq (2 + o_n(1)) \cdot n^{3/2}$

Recall that the true optimum is $\approx 1.526 \cdot n^{3/2}$ w.h.p.
Can some other algorithm do better?
In particular, how well does the Sum of Squares hierarchy do?

Our results

Theorem: SoS lower bounds for Sherrington-Kirkpatrick

For some $\delta > 0$, degree n^δ SoS cannot certify better than $(2 - o(1))n^{3/2}$

Our results

Theorem: SoS lower bounds for Sherrington-Kirkpatrick

For some $\delta > 0$, degree n^δ SoS cannot certify better than $(2 - o(1))n^{3/2}$

This vastly improves on earlier works [Mohanty et al., 2020, Kunisky and Bandeira, 2019] who showed lower bounds for degree-4 SoS

An independent work [Kunisky, 2020] obtained degree-6 lower bounds via different techniques

SoS lower bound for SK: Moving to PAP

Sherrington-Kirkpatrick: Given $W \in \mathbb{R}^{n \times n}$, determine $\max_{x \in \{-1,1\}^n} x^T W x$

SoS lower bound for SK: Moving to PAP

Sherrington-Kirkpatrick: Given $W \in \mathbb{R}^{n \times n}$, determine $\max_{x \in \{-1,1\}^n} x^T W x$

$\Updownarrow$ Take the top eigenspace
and plant a boolean vector in it
(Due to [Mohanty et al., 2020])

Planted Boolean vector: Let $p \ll n$, given a random p-dimensional subspace of R^n, determine if it contains a boolean vector $x \in \{-1, 1\}^n$

SoS lower bound for SK: Moving to PAP

Sherrington-Kirkpatrick: Given $W \in \mathbb{R}^{n \times n}$, determine $\max_{x \in \{-1,1\}^n} x^T W x$

$\Updownarrow$

Take the top eigenspace
and plant a boolean vector in it
(Due to [Mohanty et al., 2020])

Planted Boolean vector: Let $p \ll n$, given a random p-dimensional subspace of R^n, determine if it contains a boolean vector $x \in \{-1, 1\}^n$

$\Updownarrow$

Just transpose the matrix of inputs
Rename p to n and n to m

Planted Affine planes: Given vectors $d_1, \ldots, d_m \in \mathbb{R}^n$, determine if there exists a vector $v \in \{\pm\frac{1}{\sqrt{n}}\}^n$ such that $\langle v, d_u \rangle^2 = 1$ for all $u \leq m$.

SoS lower bound for SK: Moving to PAP

Reductions carry over in the SoS framework

SoS lower bound for SK: Moving to PAP

Reductions carry over in the SoS framework

Therefore, it suffices to prove SoS lower bounds for PAP for Gaussian inputs

SoS lower bound for SK: Moving to PAP

Reductions carry over in the SoS framework

Therefore, it suffices to prove SoS lower bounds for PAP for Gaussian inputs

Theorem: SoS Lower bound for Planted Affine Planes

There exists a constant $C > 0$ such that for all $\epsilon > 0$, when $m \leq n^{1.5-\epsilon}$ vectors $d_1, \ldots, d_m$ are sampled from $\mathcal{N}(0, I_n)$, w.h.p., degree-$n^{C\epsilon}$ SoS thinks the system of equations $\langle v, d_u \rangle^2 = 1$ is feasible.

SoS lower bounds for PAP: An overview

Exhibiting SoS lower bounds contains two main steps.

- Construct a candidate pseudoexpectation operator
- Show that it's feasible

SoS lower bounds for PAP: An overview

Exhibiting SoS lower bounds contains two main steps.

- Construct a candidate pseudoexpectation operator
- Show that it's feasible

For the first step, we use the technique of pseudo-calibration [Barak et al., 2016]

SoS lower bounds for PAP: An overview

Exhibiting SoS lower bounds contains two main steps.

- Construct a candidate pseudoexpectation operator
- Show that it's feasible

For the first step, we use the technique of pseudo-calibration [Barak et al., 2016]

Our main contribution lies in showing feasibility of this pseudo-expectation

SoS lower bounds for PAP: An overview

Exhibiting SoS lower bounds contains two main steps.

- Construct a candidate pseudoexpectation operator
- Show that it's feasible

For the first step, we use the technique of pseudo-calibration [Barak et al., 2016]

Our main contribution lies in showing feasibility of this pseudo-expectation

The condition $\widetilde{\mathbb{E}}[f^2] \geq 0$ for all f is usually the hardest to show

SoS lower bounds for PAP: An overview

Exhibiting SoS lower bounds contains two main steps.

- Construct a candidate pseudoexpectation operator
- Show that it's feasible

For the first step, we use the technique of pseudo-calibration [Barak et al., 2016]

Our main contribution lies in showing feasibility of this pseudo-expectation

The condition $\widetilde{\mathbb{E}}[f^2] \geq 0$ for all f is usually the hardest to show

This can be equivalently stated as showing positive-semidefiniteness w.h.p. of a large matrix Λ

- This is our main contribution

SoS lower bounds for PAP: Candidate SoS solution

Using pseudo-calibration, we obtain

$$\Lambda = \sum_{\substack{\text{shapes } \tau \\ \text{satisfying parity constraints}}} \underbrace{\frac{1}{n^{\frac{|U_\tau|+V_\tau|+|E(\tau)|}{2}}} \cdot \prod_{\textcircled{u} \in V(\tau)} h_{\deg(\textcircled{u})}(1) \cdot}_{\text{coefficient}} \underbrace{M_\tau}_{\text{graph matrix}}$$

SoS lower bounds for PAP: Candidate SoS solution

Using pseudo-calibration, we obtain

$$\Lambda = \sum_{\substack{\text{shapes } \tau \\ \text{satisfying parity constraints}}} \underbrace{\frac{1}{n^{\frac{|U_\tau|+V_\tau|+|E(\tau)|}{2}}} \cdot \prod_{\textcircled{u} \in V(\tau)} h_{\deg(\textcircled{u})}(1) \cdot}_{\text{coefficient}} \underbrace{M_\tau}_{\text{graph matrix}}$$

Here, shapes τ have two *types* of vertices - square $\boxed{i}$ and circle $\textcircled{u}$

- An edge between $\boxed{i}$ and $\textcircled{u}$ corresponds to the input variable d_{ui}
- Edges have labels corresponding to a basis element
- Generalized graph matrices introduced by [Ahn et al., 2020] who also obtain norm bounds (trace method) and show various applications

SoS lower bounds for PAP: Candidate SoS solution

Using pseudo-calibration, we obtain

$$\Lambda = \sum_{\substack{\text{shapes } \tau \\ \text{satisfying parity constraints}}} \underbrace{\frac{1}{n^{\frac{|U_\tau|+V_\tau|+|E(\tau)|}{2}}} \cdot \prod_{\textcircled{u} \in V(\tau)} h_{\deg(\textcircled{u})}(1) \cdot}_{\text{coefficient}} \underbrace{M_\tau}_{\text{graph matrix}}$$

Here, shapes τ have two *types* of vertices - square $\boxed{i}$ and circle $\textcircled{u}$

- An edge between $\boxed{i}$ and $\textcircled{u}$ corresponds to the input variable d_{ui}
- Edges have labels corresponding to a basis element
- Generalized graph matrices introduced by [Ahn et al., 2020] who also obtain norm bounds (trace method) and show various applications

In the previous section, we studied M_τ individually, but here we study their linear combinations

SoS lower bound for PAP: Identifying signal terms

First approach: Write $\Lambda = \Lambda_{sig} + \Lambda_{noise}$ and argue that

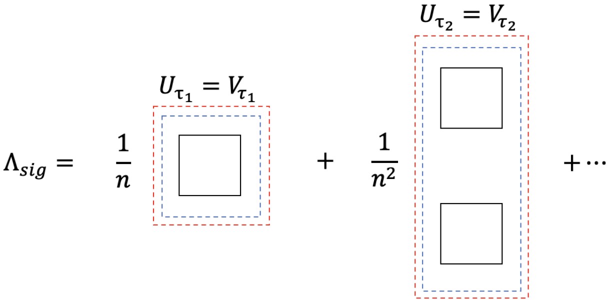
- $\Lambda_{sig} \succcurlyeq 0$
- Λ_{sig} dominates Λ_{noise} w.h.p.

SoS lower bound for PAP: Identifying signal terms

First approach: Write $\Lambda = \Lambda_{sig} + \Lambda_{noise}$ and argue that

- $\Lambda_{sig} \succ 0$
- Λ_{sig} dominates Λ_{noise} w.h.p.

A natural candidate for Λ_{sig} is "trivial" shapes corresponding to identity matrices (living in different blocks)



SoS lower bound for PAP: Charging noise terms

We hope to "charge" the rest of the terms against these signal terms

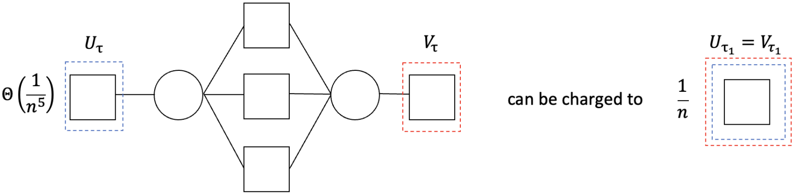
- The charging in other works including our other results, usually require nontrivial arguments that are tailored to the problem

SoS lower bound for PAP: Charging noise terms

We hope to "charge" the rest of the terms against these signal terms

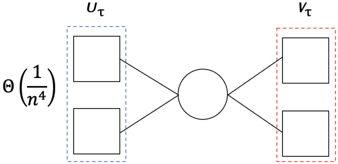
- The charging in other works including our other results, usually require nontrivial arguments that are tailored to the problem

Does work out nicely for many shapes, for example,



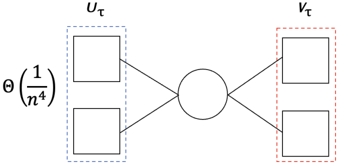
SoS lower bound for PAP: Spiders

But unfortunately, the charging fails for certain shapes, for example,



SoS lower bound for PAP: Spiders

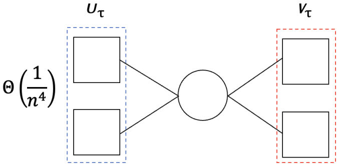
But unfortunately, the charging fails for certain shapes, for example,



Retrospectively, this failure is expected because Λ has a nontrivial kernel arising due to the constraints $\widetilde{\mathbb{E}}[\langle v, d_u\rangle^2] = 1$

SoS lower bound for PAP: Spiders

But unfortunately, the charging fails for certain shapes, for example,



Retrospectively, this failure is expected because Λ has a nontrivial kernel arising due to the constraints $\widetilde{\mathbb{E}}[\langle v, d_u \rangle^2] = 1$

We argue that all "bad" shapes must have a certain graphical substructure in them

- We call all shapes with such graphical substructure spiders

SoS lower bound for SK: Handling spiders

Our main strategy to handle spiders is to observe that they approximately lie in the kernel

SoS lower bound for SK: Handling spiders

Our main strategy to handle spiders is to observe that they approximately lie in the kernel

Proposition

Let M, A be matrices such that $MA = 0$. If $x \perp \mathrm{Null}(M)$, then $x^\intercal Mx = x^\intercal(M + A)x$

SoS lower bound for SK: Handling spiders

Our main strategy to handle spiders is to observe that they approximately lie in the kernel

Proposition

Let M, A be matrices such that $MA = 0$. If $x \perp \mathrm{Null}(M)$, then $x^\intercal M x = x^\intercal (M + A) x$

Therefore, if spiders live in the kernel, we can subtract them off

SoS lower bound for SK: Handling spiders

Our main strategy to handle spiders is to observe that they approximately lie in the kernel

Proposition

Let M, A be matrices such that $MA = 0$. If $x \perp \mathrm{Null}(M)$, then $x^\intercal Mx = x^\intercal(M + A)x$

Therefore, if spiders live in the kernel, we can subtract them off

Multiple issues arise

- Spiders only *approximately* live in the kernel, so "squashing" them may create more spiders
- We deal with them recursively and study the evolution of the coefficients, creating a web of spiders
- Charging arguments need to "remember" original coefficients

Table of Contents

Conclusion

We presented a general technique based on Matrix Efron-Stein to analyze concentration of polynomial random matrices

Conclusion

We presented a general technique based on Matrix Efron-Stein to analyze concentration of polynomial random matrices

We show high-degree SoS lower bounds for various problems

- This talk: Sherrington-Kirkpatrick Hamiltonian
- Other works: Sparse PCA, Sparse Independent Set, Tensor PCA, Planted Slightly Denser Subgraph, etc.

Conclusion

We presented a general technique based on Matrix Efron-Stein to analyze concentration of polynomial random matrices

We show high-degree SoS lower bounds for various problems

- This talk: Sherrington-Kirkpatrick Hamiltonian
- Other works: Sparse PCA, Sparse Independent Set, Tensor PCA, Planted Slightly Denser Subgraph, etc.

Our main contribution lies in the analysis of the polynomial random matrix obtained via pseudo-calibration

Open problems

1. Can we generalize our Efron-Stein framework to when the inputs are not necessarily independent?
   - Motivation: d-regular graphs instead of Erdős-Rényi graphs $\mathcal{G}_{n,d/n}$.

Open problems

1. Can we generalize our Efron-Stein framework to when the inputs are not necessarily independent?
    - Motivation: d-regular graphs instead of Erdős-Rényi graphs $\mathcal{G}_{n,d/n}$.
2. SoS lower bounds for other problems
    - Densest k-subgraph: Find the densest k-subgraph in a graph $G \sim \mathcal{G}_{n,p}$ for $p = o(1)$
        - Relates to intriguing predictions of the *log-density framework*

Open problems

1. Can we generalize our Efron-Stein framework to when the inputs are not necessarily independent?
    - Motivation: d-regular graphs instead of Erdős-Rényi graphs $\mathcal{G}_{n,d/n}$.
2. SoS lower bounds for other problems
    - Densest k-subgraph: Find the densest k-subgraph in a graph $G \sim \mathcal{G}_{n,p}$ for $p = o(1)$
        - Relates to intriguing predictions of the *log-density framework*
    - Can we unify all these lower bounds with one global theorem?
        - Closely related to the *low-degree likelihood ratio hypothesis*

Open problems

1. Can we generalize our Efron-Stein framework to when the inputs are not necessarily independent?
    - Motivation: d-regular graphs instead of Erdős-Rényi graphs $\mathcal{G}_{n,d/n}$.
2. SoS lower bounds for other problems
    - Densest k-subgraph: Find the densest k-subgraph in a graph $G \sim \mathcal{G}_{n,p}$ for $p = o(1)$
        - Relates to intriguing predictions of the *log-density framework*
    - Can we unify all these lower bounds with one global theorem?
        - Closely related to the *low-degree likelihood ratio hypothesis*
3. How does SoS compare with other algorithmic techniques such as low-degree polynomials or statistical query algorithms?

Open problems

1. Can we generalize our Efron-Stein framework to when the inputs are not necessarily independent?
    - Motivation: d-regular graphs instead of Erdős-Rényi graphs $\mathcal{G}_{n,d/n}$.
2. SoS lower bounds for other problems
    - Densest k-subgraph: Find the densest k-subgraph in a graph $G \sim \mathcal{G}_{n,p}$ for $p = o(1)$
        - Relates to intriguing predictions of the *log-density framework*
    - Can we unify all these lower bounds with one global theorem?
        - Closely related to the *low-degree likelihood ratio hypothesis*
3. How does SoS compare with other algorithmic techniques such as low-degree polynomials or statistical query algorithms?
4. In current SoS lower bounds, can we improve the degree of SoS from n^{ϵ} to $\Omega(n)$?

Open problems

1. Can we generalize our Efron-Stein framework to when the inputs are not necessarily independent?
    - Motivation: d-regular graphs instead of Erdős-Rényi graphs $\mathcal{G}_{n,d/n}$.
2. SoS lower bounds for other problems
    - Densest k-subgraph: Find the densest k-subgraph in a graph $G \sim \mathcal{G}_{n,p}$ for $p = o(1)$
        - Relates to intriguing predictions of the *log-density framework*
    - Can we unify all these lower bounds with one global theorem?
        - Closely related to the *low-degree likelihood ratio hypothesis*
3. How does SoS compare with other algorithmic techniques such as low-degree polynomials or statistical query algorithms?
4. In current SoS lower bounds, can we improve the degree of SoS from n^{ϵ} to $\Omega(n)$?
5. Can we simplify the current proofs?

Thank You

(No applause please)

Adamczak, R. and Wolff, P. (2015).
Concentration inequalities for non-lipschitz functions with bounded derivatives of higher order.
Probability Theory and Related Fields, 162(3):531–586.

Ahn, K., Medarametla, D., and Potechin, A. (2020).
Graph matrices: Norm bounds and applications.
abs/1604.03423.

Arora, S., Rao, S., and Vazirani, U. (2004).
Expander flows and a $\sqrt{\log n}$-approximation to sparsest cut.
In *Proceedings of the 36th ACM Symposium on Theory of Computing*.

Barak, B., Brandão, F. G. S. L., Harrow, A. W., Kelner, J. A., Steurer, D., and Zhou, Y. (2012).
Hypercontractivity, sum-of-squares proofs, and their applications.
CoRR, abs/1205.4484.

Barak, B., Hopkins, S. B., Kelner, J., Kothari, P., Moitra, A., and Potechin, A. (2016).

A nearly tight sum-of-squares lower bound for the planted clique problem.
In *Proceedings of the 57th IEEE Symposium on Foundations of Computer Science*, pages 428–437.

Boucheron, S., Bousquet, O., Lugosi, G., and Massart, P. (2005).
Moment inequalities for functions of independent random variables.
The Annals of Probability, 33(2):514 – 560.

Chatterjee, S. (2005).
Concentration inequalities with exchangeable pairs.
Stanford University.

Crisanti, A. and Rizzo, T. (2002).
Analysis of the ∞-replica symmetry breaking solution of the sherrington-kirkpatrick model.
Physical Review E, 65(4):046137.

Ge, R. and Ma, T. (2015).
Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms.

arXiv preprint arXiv:1504.05287.

Goemans, M. and Williamson, D. (1995).
Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming.
Journal of the ACM, 42(6):1115–1145.
Preliminary version in *Proc. of STOC'94*.

Grigoriev, D. (2001).
Complexity of positivstellensatz proofs for the knapsack.
computational complexity, 10(2):139–154.

Hopkins, S. B., Shi, J., and Steurer, D. (2015).
Tensor principal component analysis via sum-of-square proofs.
In *Conference on Learning Theory*, pages 956–1006. PMLR.

Kim, J. H. and Vu, V. H. (2000).
Concentration of multivariate polynomials and its applications.
Combinatorica, 20(3):417–434.

Kunisky, D. (2020).

Positivity-preserving extensions of sum-of-squares pseudomoments over the hypercube.
arXiv preprint arXiv:2009.07269.

Kunisky, D. and Bandeira, A. S. (2019).
A tight degree 4 sum-of-squares lower bound for the sherrington-kirkpatrick hamiltonian.
abs/1907.11686.

Lasserre, J. B. (2001).
Global optimization with polynomials and the problem of moments.
SIAM Journal on optimization, 11(3):796–817.

Latała, R. (2006).
Estimates of moments and tails of gaussian chaoses.
The Annals of Probability, 34(6):2315–2331.

Medarametla, D. and Potechin, A. (2016).
Bounds on the norms of uniform low degree graph matrices.
RANDOM.

Mohanty, S., Raghavendra, P., and Xu, J. (2020).

Lifting sum-of-squares lower bounds: degree-2 to degree-4.
In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 840–853.

Moitra, A. and Wein, A. S. (2019).
Spectral methods from tensor networks.
In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 926–937.

Montanari, A. (2019).
Optimization of the sherrington-kirkpatrick hamiltonian.
In *Proceedings of the 60th IEEE Symposium on Foundations of Computer Science*, pages 1417–1433.

Panchenko, D. (2014).
The parisi formula for mixed p-spin models.
Ann. Probab., 42(3):946–958.

Parisi, G. (1979).
Infinite number of order parameters for spin-glasses.
Phys. Rev. Lett., 43:1754–1756.

Parrilo, P. A. (2000).
Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization.
PhD thesis, California Institute of Technology.

Paulin, D., Mackey, L., and Tropp, J. A. (2016).
Efron–stein inequalities for random matrices.
Ann. Probab., 44(5):3431–3473.

Schramm, T. and Steurer, D. (2017).
Fast and robust tensor decomposition with applications to dictionary learning.
In *Conference on Learning Theory*, pages 1760–1793. PMLR.

Schudy, W. and Sviridenko, M. (2011).
Bernstein-like concentration and moment inequalities for polynomials of independent random variables: multilinear case.
arXiv preprint arXiv:1109.5193.

Stein, C. (1972).

A bound for the error in the normal approximation to the distribution of a sum of dependent random variables.
In *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: Probability theory*, pages 583–602. University of California Press.

Talagrand, M. (2006).
The parisi formula.
Annals of mathematics, pages 221–263.

Tulsiani, M. (2009).
CSP gaps and reductions in the Lasserre hierarchy.
In *Proceedings of the 41st ACM Symposium on Theory of Computing*.