Planted Clique, Sum-of-Squares and Pseudo-Calibration

Ankur Moitra (MIT)

joint work with Boaz Barak, Sam Hopkins, Jon Kelner, Pravesh Kothari and Aaron Potechin
PLANTED CLIQUE

Introduced by [Jerrum, ’92], [Kucera, ’95]:
PLANTED CLIQUE

Introduced by [Jerrum, ‘92], [Kucera, ’95]:

**Step #1:** Generate E-R random graph $G(n, \frac{1}{2})$
PLANTED CLIQUE

Introduced by [Jerrum, ‘92], [Kucera, ’95]:

**Step #1:** Generate E-R random graph $G(n, \frac{1}{2})$

**Step #2:** Add a clique on random set of $\omega$ vertices
PLANTED CLIQUE

Introduced by [Jerrum, ‘92], [Kucera, ’95]:

**Step #1:** Generate E-R random graph $G(n, \frac{1}{2})$

**Step #2:** Add a clique on random set of $\omega$ vertices
PLANTED CLIQUE

Introduced by [Jerrum, ‘92], [Kucera, ’95]:

**Step #1:** Generate E-R random graph $G(n, \frac{1}{2})$

**Step #2:** Add a clique on random set of $\omega$ vertices

Can we find the planted clique?

And how large does $\omega$ need to be?
Quasi-polynomial time:

**Fact:** There is an $n^{O(\log n)}$-time algorithm (brute-force) that can find planted cliques of size $\omega \geq C \log n$, for any $C > 2$
Quasi-polynomial time:

Fact: There is an $n^{O(\log n)}$-time algorithm (brute-force) that can find planted cliques of size $\omega \geq C \log n$, for any $C > 2$

Polynomial time:

Fact: There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n \log n}$ (degree counting)
Quasi-polynomial time:

**Fact:** There is an $n^{O(\log n)}$-time algorithm (brute-force) that can find planted cliques of size $\omega \geq C \log n$, for any $C > 2$.

Polynomial time:

**Fact:** There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n \log n}$ (degree counting).

**Theorem [Alon, Krivelevich, Sudakov ‘98]:** There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n}$ (spectral).
**Quasi-polynomial time:**

**Fact:** There is an $n^{O(\log n)}$-time algorithm (brute-force) that can find planted cliques of size $\omega \geq C \log n$, for any $C > 2$

---

**Polynomial time:**

**Fact:** There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n \log n}$ (degree counting)

**Theorem [Alon, Krivelevich, Sudakov ‘98]:** There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n}$ (spectral)

**Theorem [Deshpande, Montanari ‘13]:** There is a nearly linear time algorithm that succeeds (whp) for $\omega \geq \sqrt{n/e}$
APPLICATIONS OF PLANTED CLIQUE

Planted Clique (and variants) are basic problems in average-case analysis, many applications:
APPLICATIONS OF PLANTED CLIQUE

Planted Clique (and variants) are basic problems in average-case analysis, many applications:

- Discovering motifs in biological networks [Milo et al ‘02]
- Computing the best Nash Equilibrium [HK ‘11], [ABC ‘13]
- Property testing [Alon et al ‘07]
- Sparse PCA [Berthet, Rigollet ‘13]
- Compressed sensing [Koiran, Zouzias ‘14]
- Cryptography [Juels, Peinado ‘00], [Applebaum et al ‘10]
- Mathematical finance [Arora et al ‘10]
LOWER BOUNDS?

Is it *actually* hard to find $n^{1/2-\varepsilon}$-sized planted cliques?
LOWER BOUNDS?

Is it actually hard to find $n^{1/2-\varepsilon}$-sized planted cliques?

Complexity-theoretic reasons lower bound are unlikely to be based on $P$ vs. $NP$

e.g. [Feigenbaum, Fortnow ’93], [Bogdanov, Trevisan ’06]
LOWER BOUNDS?

Is it *actually* hard to find $n^{1/2-\varepsilon}$-sized planted cliques?

Complexity-theoretic reasons lower bound are unlikely to be based on **P vs. NP**

e.g. [Feigenbaum, Fortnow ’93], [Bogdanov, Trevisan ’06]

Our best evidence seems to come from **hierarchies**...
OUTLINE

Part I: Introduction

• Planted Clique and its Applications
• The Sum-of-Squares Hierarchy
• Our Results

Part II: Fooling SOS

• The Meka-Potechin-Wigderson Moments
• Kelner’s Polynomial, and Corrections at $d = 4$
• Pseudo-Calibration and Fourier Analysis
• Symbolic Factorization and Intersection Terms
OUTLINE

Part I: Introduction

• Planted Clique and its Applications
• The Sum-of-Squares Hierarchy
• Our Results

Part II: Fooling SOS

• The Meka-Potechin-Wigderson Moments
• Kelner’s Polynomial, and Corrections at $d = 4$
• Pseudo-Calibration and Fourier Analysis
• Symbolic Factorization and Intersection Terms
SUM-OF-SQUARES HIERARCHY

Powerful hierarchy of semidefinite programs, introduced by [Shor ‘87], [Nesterov ‘00], [Parrilo ‘00], [Lasserre ‘01]
SUM-OF-SQUARES HIERARCHY

Powerful hierarchy of semidefinite programs, introduced by [Shor ‘87], [Nesterov ‘00], [Parrilo ‘00], [Lasserre ‘01]

**Goal:** Find operator that behaves like the expectation over a distribution on solutions

\[ \tilde{E} : \mathcal{P}_{\leq d}^{n} \rightarrow \mathbb{R} \]

degree \( \leq d \) polynomials in \( n \) variables

Called a **Pseudo-expectation**
Constraints on the pseudo-expectation:

(1) $\tilde{\mathbb{E}}$ is linear

(2) $\tilde{\mathbb{E}}[1] = 1$

(3) $\tilde{\mathbb{E}}[\rho^2] \geq 0$

for all $\text{deg}(\rho) \leq d/2$

**general**
Constraints on the pseudo-expectation:

1. \( \tilde{\mathbb{E}} \) is linear
2. \( \tilde{\mathbb{E}}[1] = 1 \)
3. \( \tilde{\mathbb{E}}[p^2] \geq 0 \) for all \( \deg(p) \leq d/2 \)
4. \( \tilde{\mathbb{E}}[x_i^2 p] = \tilde{\mathbb{E}}[x_i p] \) (booleanity)
Constraints on the pseudo-expectation:

1. $\tilde{\mathbb{E}}$ is linear

2. $\tilde{\mathbb{E}}[1] = 1$

3. $\tilde{\mathbb{E}}[\rho^2] \geq 0$

   for all $\deg(p) \leq d/2$

4. $\tilde{\mathbb{E}}[x_i^2p] = \tilde{\mathbb{E}}[x_ip]$

5. $\tilde{\mathbb{E}}[\sum x_i] = \omega$

   (clique size)
Constraints on the pseudo-expectation:

(1) \( \widetilde{\mathbb{E}} \) is linear

(2) \( \widetilde{\mathbb{E}}[1] = 1 \)

(3) \( \widetilde{\mathbb{E}}[p^2] \geq 0 \)

for all \( \text{deg}(p) \leq d/2 \)

(4) \( \widetilde{\mathbb{E}}[x_i^2p] = \widetilde{\mathbb{E}}[x_ip] \)

(5) \( \widetilde{\mathbb{E}}[\sum x_i] = \omega \)

(6) \( \widetilde{\mathbb{E}}[x_ix_jp] = 0 \)

for all \((i,j)\) not an edge

\(\text{(clique constraints)}\)
Constraints on the pseudo-expectation:

1. \( \tilde{E} \) is linear
2. \( \tilde{E}[1] = 1 \)
3. \( \tilde{E}[p^2] \geq 0 \) for all \( \deg(p) \leq d/2 \)
4. \( \tilde{E}[x_i^2p] = \tilde{E}[x_ip] \)
5. \( \tilde{E}[\sum x_i] = \omega \)
6. \( \tilde{E}[x_ix_jp] = 0 \) for all \((i,j)\) not an edge

**general**

**specific to planted clique**
Constraints on the pseudo-expectation:

1. $\mathbb{E}$ is linear
2. $\mathbb{E}[1] = 1$
3. $\mathbb{E}[p^2] \geq 0$

for all $\deg(p) \leq d/2$

4. $\mathbb{E}[x_i^2p] = \mathbb{E}[x_ip]$
5. $\mathbb{E}\left[\sum x_i\right] = \omega$
6. $\mathbb{E}[x_ix_jp] = 0$

for all $(i, j)$ not an edge

E.g. if $a_1, a_2, \ldots, a_n$ is the indicator vector of an $\omega$-sized clique

$$\mathbb{E}[p(x_1, x_2, \ldots, x_n)] = p(a_1, a_2, \ldots, a_n)$$

meets (1) – (6)
Constraints on the pseudo-expectation:

(1) \( \widetilde{E} \) is linear

(2) \( \widetilde{E}[1] = 1 \)

(3) \( \widetilde{E}[p^2] \geq 0 \) for all \( \deg(p) \leq d/2 \)

(4) \( \widetilde{E}[x_i^2 p] = \widetilde{E}[x_i p] \)

(5) \( \widetilde{E}[\sum x_i] = \omega \)

(6) \( \widetilde{E}[x_i x_j p] = 0 \) for all \((i,j)\) not an edge

There is an \( n^{O(d)} \)-time algorithm for finding such an operator, if it exists

Called the level d **Sum-of-Squares Algorithm**
• strengthens **Sherali-Adams, Lovasz-Schrijver, LS+**

• breaks integrality gaps for other hierarchies [Barak et al, ‘12]

• highly successful convex relaxation
  - sparsest cut [ARV ‘04]
  - unique games [ABS ‘10], [BRS ‘12], [GS ‘12]

• optimal among all poly. sized SDPs for random CSPs [LRS ‘15]

• best known algorithm for several **average-case** problems
  - planted sparse vector, dictionary learning [BKS ‘14, ‘15]
  - noisy tensor completion [BM ‘15], tensor PCA [HSS ‘15]
• strengthens **Sherali-Adams, Lovasz-Schrijver, LS+**

• breaks integrality gaps for other hierarchies [Barak et al, ‘12]

• highly successful convex relaxation
  
  sparsest cut [ARV ’04]
  
  unique games [ABS ’10], [BRS ’12], [GS ’12]

• optimal among all poly. sized SDPs for random CSPs [LRS ‘15]

• best known algorithm for several **average-case** problems
  
  planted sparse vector, dictionary learning [BKS ’14, ’15]
  
  noisy tensor completion [BM ’15], tensor PCA [HSS ’15]

Can it find \( n^\varepsilon \)-sized planted cliques in polynomial time?
OUTLINE

Part I: Introduction

• Planted Clique and its Applications
• The Sum-of-Squares Hierarchy
• Our Results

Part II: Fooling SOS

• The Meka-Potechin-Wigderson Moments
• Kelner’s Polynomial, and Corrections at $d = 4$
• Pseudo-Calibration and Fourier Analysis
• Symbolic Factorization and Intersection Terms
OUTLINE

Part I: Introduction

• Planted Clique and its Applications
• The Sum-of-Squares Hierarchy
• Our Results

Part II: Fooling SOS

• The Meka-Potechin-Wigderson Moments
• Kelner’s Polynomial, and Corrections at $d = 4$
• Pseudo-Calibration and Fourier Analysis
• Symbolic Factorization and Intersection Terms
OUR RESULTS

We show a nearly optimal lower bound against SoS, for the planted clique problem:

Theorem [Barak, Hopkins, Kelner, Kothari, Moitra, Potechin]:
The integrality gap of the level $d$ Sum-of-Squares hierarchy is

$$n^{\frac{1}{2} - c \sqrt{d/ \log n}}$$

for some constant $c > 0$

For any $d = o(\log n)$, the integrality gap is $n^{1/2 - o(1)}$
OUR RESULTS

We show a nearly optimal lower bound against SoS, for the planted clique problem:

**Theorem [Barak, Hopkins, Kelner, Kothari, Moitra, Potechin]:**
The integrality gap of the level $d$ Sum-of-Squares hierarchy is

$$n^{\frac{1}{2} - c \sqrt{d / \log n}}$$

for some constant $c > 0$

For any $d = o(\log n)$, the integrality gap is $n^{1/2-o(1)}$

Improves upon [Meka, Potechin, Wigderson ‘14], [Deshpande Montanari ‘15], [Hopkins, Kothari, Potechin, Raghavendra, Scrhamm ‘16]
OUR RESULTS

Our Approach: **Pseudo-calibration**

New insights into what makes SoS powerful, and how to fool it
OUR RESULTS

Our Approach: **Pseudo-calibration**

New insights into what makes SoS powerful, and how to fool it

When our *recipe* fails, does it immediately yield algorithms?
OUTLINE

Part I: Introduction
• Planted Clique and its Applications
• The Sum-of-Squares Hierarchy
• Our Results

Part II: Fooling SOS
• The Meka-Potechin-Wigderson Moments
• Kelner’s Polynomial, and Corrections at $d = 4$
• Pseudo-Calibration and Fourier Analysis
• Symbolic Factorization and Intersection Terms
OUTLINE

Part I: Introduction

• Planted Clique and its Applications
• The Sum-of-Squares Hierarchy
• Our Results

Part II: Fooling SOS

• The Meka-Potechin-Wigderson Moments
• Kelner’s Polynomial, and Corrections at $d = 4$
• Pseudo-Calibration and Fourier Analysis
• Symbolic Factorization and Intersection Terms
How can we fool the SoS algorithm into thinking there is a $n^{1/2-o(1)}$ sized clique in $G(n, \frac{1}{2})$?
PSEUDO-MOMENTS

How can we fool the SoS algorithm into thinking there is a $n^{1/2-o(1)}$ sized clique in $G(n, ½)$?

**Usual Approach:** Adapt integrality gaps from weaker hierarchies
How can we fool the SoS algorithm into thinking there is a $n^{1/2-o(1)}$ sized clique in $G(n, \frac{1}{2})$?

**Usual Approach:** Adapt integrality gaps from weaker hierarchies

This works for random CSPs
How can we fool the SoS algorithm into thinking there is a $n^{1/2-o(1)}$ sized clique in $G(n, \frac{1}{2})$?

**Usual Approach:** Adapt integrality gaps from weaker hierarchies

This works for random CSPs

**Theorem [Feige, Krauthgamer ‘03]:** The integrality gap of the level $d$ LS+ hierarchy is

$$\sqrt{\frac{n}{2^d}}$$
Theorem [Meka, Potechin, Wigderson ‘14]: The integrality gap of the level $d$ Sum-of-Squares hierarchy is

\[ \eta^{1/d - o(1)} \]
Theorem [Meka, Potechin, Wigderson ‘14]: The integrality gap of the level $d$ Sum-of-Squares hierarchy is

$$\eta^{1/d - o(1)}$$

In particular, set:

$$\widetilde{E}_{MPW} \left[ \prod_{i \in A} x_i \right] = 2^{|A|/2} \left( \frac{\omega}{n} \right)^{|A|}$$

if $A$ is clique, zero otherwise.
**Theorem [Meka, Potechin, Wigderson ‘14]:** The integrality gap of the level \(d\) Sum-of-Squares hierarchy is

\[ n^{1/d - o(1)} \]

In particular, set:

\[
\mathbb{E}_{MPW} \left[ \prod_{i \in A} x_i \right] = 2^{\frac{|A|}{2}} \left( \frac{\omega}{n} \right)^{|A|}
\]

if \(A\) is clique, zero otherwise. Extend by linearity to all \(\deg(p) \leq d\).
Theorem [Meka, Potechin, Wigderson ‘14]: The integrality gap of the level $d$ Sum-of-Squares hierarchy is

$$n^{1/d-o(1)}$$

In particular, set:

$$\widetilde{\mathbb{E}}_{MVPW} \left[ \prod_{i \in A} x_i \right] = 2^{\left( \frac{|A|}{2} \right)} \left( \frac{\omega}{n} \right)^{|A|}$$

if $A$ is clique, zero otherwise. Extend by linearity to all $\deg(p) \leq d$

**Approach:** Spectral bounds on **locally random matrices**
Theorem [Meka, Potechin, Wigderson ‘14]: The integrality gap of the level $d$ Sum-of-Squares hierarchy is

$$n^{1/d - o(1)}$$
Theorem [Meka, Potechin, Wigderson ‘14]: The integrality gap of the level $d$ Sum-of-Squares hierarchy is

$$n^{1/d-o(1)}$$

Improved analysis due to [Deshpande, Montanari ’15], for $d = 4$

$$n^{1/3-o(1)}$$
Theorem [Meka, Potechin, Wigderson ‘14]: The integrality gap of the level $d$ Sum-of-Squares hierarchy is
\[ n^{1/d - o(1)} \]

Improved analysis due to [Deshpande, Montanari ’15], for $d = 4$
\[ n^{1/3 - o(1)} \]

And due to [Hopkins, Kothari, Potechin ’16] for any $d$
\[ n^{1/([d/2] + 1) - o(1)} \]
**Theorem [Meka, Potechin, Wigderson ’14]:** The integrality gap of the level $d$ Sum-of-Squares hierarchy is

$$\eta^{1/d-o(1)}$$

Improved analysis due to [Deshpande, Montanari ’15], for $d = 4$

$$\eta^{1/3-o(1)}$$

And due to [Hopkins, Kothari, Potechin ’16] for any $d$

$$\eta^{1/([d/2]+1)-o(1)}$$

But these bounds are **tight** (for these moments)
OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner’s Polynomial, and Corrections at $d = 4$
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms
OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner’s Polynomial, and Corrections at \( d = 4 \)
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms
KELNER’S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?
KELNER’S POLYNOMIAL

Do the MPW moments work beyond $n^{1/([d/2] + 1)}$?

Set $G_{i,j} = \begin{cases} 
+1 & \text{if } (i,j) \text{ an edge} \\
-1 & \text{else} 
\end{cases}$

Then $P_{G,i} = \left( \sum_j G_{i,j} x_j \right)^\ell$
KELNER’S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

Set $G_{i,j} = \begin{cases} 
+1 & \text{if } (i,j) \text{ an edge} \\
-1 & \text{else} 
\end{cases}$

$$P_{G,i} = \left( \sum_j G_{i,j} x_j \right)^\ell$$

If there is an $\omega$-sized planted clique:

$$\mathbb{E}[P_{G,i}^2] \geq \left( \frac{\omega}{n} \right) \omega^{2\ell}$$
KELNER’S POLYNOMIAL

Do the MPW moments work beyond $n^{1/\lceil d/2 \rceil + 1}$?

Set $G_{i,j} = \begin{cases} +1 & \text{if } (i,j) \text{ an edge} \\ -1 & \text{else} \end{cases}$

$P_{G,i} = \left( \sum_j G_{i,j} x_j \right)^\ell$

If there is an $\omega$-sized planted clique:

$$\mathbb{E}[P_{G,i}^2] \geq \left( \frac{\omega}{n} \right) \omega^{2\ell}$$

But if $G$ is sampled from $G(n, \frac{1}{2})$:

$$\mathbb{E}[\mathbb{E}_{MPW}[P_{G,i}^2]] \leq (n^\ell) \left( \frac{\omega}{n} \right)^\ell = \omega^\ell$$
KELNER’S POLYNOMIAL

Do the MPW moments work beyond $n^{1/\left(\lceil d/2 \rceil + 1 \right)}$?

Set $G_{i,j} = \begin{cases} +1 & \text{if (i,j) an edge} \\ -1 & \text{else} \end{cases}$

$$P_{G,i} = \left( \sum_j G_{i,j} x_j \right)^\ell$$

If there is an $\omega$-sized planted clique:

$$\mathbb{E}[P^2_{G,i}] \geq \left( \frac{\omega}{n} \right)^{2\ell}$$

But if $G$ is sampled from $G(n, \frac{1}{2})$:

$$\mathbb{E}[\mathbb{E}_{MPW}[P^2_{G,i}]] \leq (n^\ell) \left( \frac{\omega}{n} \right)^\ell = \omega^\ell$$

Need: $\omega \leq n^{1/(\ell+1)} = n^{1/(d/2+1)}$ otherwise something is wrong
KELNER’S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?
KELNER’S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

This example can be used to find a squared polynomial whose pseudo-expectation is negative for $\omega > n^{1/(\lceil d/2 \rceil + 1)}$

$$\mathbb{E}_{MPW}[P^2] < 0$$
Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

This example can be used to find a squared polynomial whose pseudo-expectation is negative for $\omega > n^{1/(\lceil d/2 \rceil + 1)}$

$$\hat{\mathbb{E}}_{MPW}[P^2] < 0$$

**Intuition:** A good pseudo-expectation attempts to hide info about what vertices participate in the planted clique
KELNER’S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

This example can be used to find a squared polynomial whose pseudo-expectation is negative for $\omega > n^{1/(\lceil d/2 \rceil + 1)}$

$$\tilde{E}_{MPW}[P^2] < 0$$

**Intuition:** A good pseudo-expectation attempts to hide info about what vertices participate in the planted clique

But vertices with a **standard deviation higher degree**, should be a constant factor more likely to be in the p.c. (soft constraint)
FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at $d = 4$

Theorem [Hopkins et al ’16], [Raghavendr, Schramm ‘16]:
The integrality gap of the level 4 Sum-of-Squares hierarchy is

$$n^{1/2 - o(1)}$$
FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at $d = 4$

**Theorem [Hopkins et al ’16], [Raghavendra, Schramm ‘16]:**
The integrality gap of the level 4 Sum-of-Squares hierarchy is

$$n^{1/2 - o(1)}$$

**Approach:** Add an explicit correction term of fix all $P_{G,i}$’s, even more dependent random matrix theory
FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at \( d = 4 \)

**Theorem [Hopkins et al ’16], [Raghavendra, Schramm ‘16]:** The integrality gap of the level 4 Sum-of-Squares hierarchy is

\[
\eta^{1/2-o(1)}
\]

**Approach:** Add an explicit correction term of fix all \( P_{G,i} \)'s, even more dependent random matrix theory

**Is there a fix for higher degrees?**
FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at \( d = 4 \)

**Theorem [Hopkins et al ’16], [Raghavendra, Schramm ‘16]:** The integrality gap of the level 4 Sum-of-Squares hierarchy is

\[
\eta^{1/2 - o(1)}
\]

**Approach:** Add an explicit correction term of fix all \( P_{G,i} \)'s, even more dependent random matrix theory

Is there a fix for higher degrees?

It turns out for \( d = 6 \), even the fixes need fixes, and on and on...
FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at $d = 4$

Theorem [Hopkins et al ’16], [Raghavendra, Schramm ‘16]: The integrality gap of the level 4 Sum-of-Squares hierarchy is

$$n^{1/2-o(1)}$$

Approach: Add an explicit correction term of fix all $P_{G,i}$’s, even more dependent random matrix theory

Is there a fix for higher degrees?

It turns out for $d = 6$, even the fixes need fixes, and on and on...
Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner’s Polynomial, and Corrections at $d = 4$
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms
OUTLINE

Part I: Introduction
- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS
- The Meka-Potechin-Wigderson Moments
- Kelner’s Polynomial, and Corrections at $d = 4$
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms
PSEUDO-CALIBRATION

Can we find pseudo-moments that satisfy the following:

$$
\mathbb{E}[\mathbb{E}[f(G, x)]] = \mathbb{E}[f(G, x)]
$$

for all *simple* functions $f$?
Can we find pseudo-moments that satisfy the following:

\[
\mathbb{E}\left[\mathbb{E}[f(G, x)]\right] = \mathbb{E}[f(G, x)]
\]

for all polynomials \( f \) that are low-degree in \( G_{i,j} \)'s and \( x_i \)'s?
Consider the pseudo-expectation of some monomial:

\[ \hat{E}[x_A] : G \to \mathbb{R}, \text{ and let } \chi_T(G) = \prod_{(i,j) \in T} G_{i,j} \]
Consider the pseudo-expectation of some monomial:

$$\widehat{\mathbb{E}}[x_A] : G \to \mathbb{R}, \text{ and let } \chi_T(G) = \prod_{(i,j) \in T} G_{i,j}$$

We can write any such function in terms of its **Fourier expansion**

$$\widehat{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\mathbb{E}}[x_A](T) \chi_T(G)$$
Consider the pseudo-expectation of some monomial:

\[ \hat{\mathbb{E}}[x_A] : G \to \mathbb{R}, \text{ and let } \chi_T(G) = \prod_{(i,j) \in T} G_{i,j} \]

We can write any such function in terms of its Fourier expansion

\[ \hat{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \hat{\mathbb{E}}[x_A](T) \chi_T(G) \]

How should we set the Fourier coefficients?
The Fourier coefficients are chosen for us, by pseudo-calibration
The Fourier coefficients are chosen for us, by pseudo-calibration.

Utilizing the expression

\[ \widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq ([n]/2)} \widetilde{\mathbb{E}}[x_A](T)\chi_T(G) \]

we can calculate:

\[ \mathbb{E}[\widetilde{\mathbb{E}}[x_A\chi_T(G)]] \]
Utilizing the expression

\[ \tilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \tilde{\mathbb{E}}[x_A](T) \chi_T(G) \]

we can calculate:

\[ \mathbb{E}[\tilde{\mathbb{E}}[x_A] \chi_T(G)] \] (by linearity)
Utilizing the expression

\[ \tilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{n}{2}} \tilde{\mathbb{E}}[x_A](T) \chi_T(G) \]

we can calculate:

\[ \mathbb{E}[\tilde{\mathbb{E}}[x_A] \chi_T(G)] = \sum_{T' \subseteq \binom{n}{2}} \tilde{\mathbb{E}}[x_A](T') \mathbb{E}[\chi_T(G) \chi_{T'}(G)] \]
The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

\[ \tilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \mathbb{E}[x_A](T) \chi_T(G) \]

we can calculate:

\[ \mathbb{E}[\tilde{\mathbb{E}}[x_A] \chi_T(G)] = \sum_{T' \subseteq \binom{[n]}{2}} \tilde{\mathbb{E}}[x_A](T') \mathbb{E}[\chi_T(G) \chi_{T'}(G)] \]

\[ = \begin{cases} +1 & \text{if } T = T' \\ 0 & \text{else} \end{cases} \]
Utilizing the expression

\[ \tilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \tilde{\mathbb{E}}[x_A](T) \chi_T(G) \]

we can calculate:

\[ \mathbb{E}[\tilde{\mathbb{E}}[x_A \chi_T(G)]] = \tilde{\mathbb{E}}[x_A](T) \]
The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

$$\tilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \left( \begin{array}{c} n \\ 2 \end{array} \right)} \tilde{\mathbb{E}}[x_A](T) \chi_T(G)$$

we can calculate:

$$\mathbb{E}\left[ \tilde{\mathbb{E}}[x_A \chi_T(G)] \right] = \tilde{\mathbb{E}}[x_A](T)$$

$$\overset{\triangle}{=} \mathbb{E}[x_A \chi_T(G)]$$

pseudo-calibration
Utilizing the expression

$$\tilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \tilde{\mathbb{E}}[x_A](T) \chi_T(G)$$

we can calculate:

$$\mathbb{E}[\tilde{\mathbb{E}}[x_A x_T(G)]] = \tilde{\mathbb{E}}[x_A](T)$$

pseudo-calibration

$$\triangleq \mathbb{E}[x_A x_T(G)] = \left(\frac{\omega}{n}\right)^{|V(T)\cup A|}$$

vertices of $T$
The Fourier coefficients are chosen for us, by pseudo-calibration.

Utilizing the expression

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \{\lfloor n \rfloor\}} \widetilde{\mathbb{E}}[x_A](T) \chi_T(G)$$

we can calculate:

$$\mathbb{E}[\widetilde{\mathbb{E}}[x_A \chi_T(G)]] = \widetilde{\mathbb{E}}[x_A](T)$$

$$\triangleq \mathbb{E}[x_A \chi_T(G)] = \left(\frac{\omega}{n}\right)^{|V(T) \cup A|}$$

It turns out, we need to truncate but at what degree?
TRUNCATION

Our pseudo-moments are:

$$
\tilde{E}[x_A] = \sum_{T \subseteq \binom{[n]}{2}} \left( \frac{\omega}{n} \right)^{|V(T) \cup A|} \chi_T(G)
$$

where $|V(T) \cup A| \leq \tau$
Our pseudo-moments are:

$$\tilde{\mathbb{E}}[x_A] = \sum_{T \subseteq \binom{[n]}{2}} \binom{\omega}{n} |V(T) \cup A| \chi_T(G)$$

$$|V(T) \cup A| \leq \tau$$

**Lemma:** With high probability,

$$|\tilde{\mathbb{E}}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left( \frac{\omega}{\sqrt{n}} \right)^t$$
TRUNCATION

Our pseudo-moments are:

\[
\tilde{\mathbb{E}}[x_A] = \sum_{T \subseteq \binom{[n]}{2}} \left( \frac{\omega}{n} \right)^{|V(T) \cup A|} \chi_T(G)
\]

\[
|V(T) \cup A| \leq \tau
\]

**Lemma:** With high probability,

\[
|\tilde{\mathbb{E}}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left( \frac{\omega}{\sqrt{n}} \right)^t
\]

(1) This is why we need to truncate
Our pseudo-moments are:

\[ \tilde{E}[x_A] = \sum_{T \subseteq ([n]_2)} \left( \frac{\omega}{n} \right)^{|V(T) \cup A|} \chi_T(G) \]

\[ |V(T) \cup A| \leq \tau \]

**Lemma:** With high probability, 

\[ |\tilde{E}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left( \frac{\omega}{\sqrt{n}} \right)^t \]

\[ n^{-\Omega(\epsilon)} \]

(2) Is small enough for any \( \omega \leq n^{1/2 - \epsilon} \) for \( \tau \leq \frac{\epsilon}{2 \log n} \)
Our pseudo-moments are:

\[
\begin{align*}
\widetilde{\mathbb{E}}[x_A] &= \sum_{T \subseteq [n]} \left( \frac{\omega}{n} \right)^{|V(T) \cup A|} \chi_T(G) \\
&\quad \text{such that } |V(T) \cup A| \leq \tau
\end{align*}
\]

**Lemma:** With high probability,

\[
|\widetilde{\mathbb{E}}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left( \frac{\omega}{\sqrt{n}} \right)^t
\]

(3) Can always renormalize pseudo-expectation so \(\widetilde{\mathbb{E}}[1] = 1\)
Our pseudo-moments are:

\[ \widetilde{E}[x_A] = \sum_{T \subseteq \binom{[n]}{2}} \left( \frac{\omega}{n} \right)^{|V(T) \cup A|} \chi_T(G) \]

\[ |V(T) \cup A| \leq \tau \]

**Lemma:** With high probability,

\[ |\widetilde{E}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left( \frac{\omega}{\sqrt{n}} \right)^t \]

(4) Similar bound holds (again by standard concentration) for

\[ \widetilde{E}\left[ \sum_i x_i \right] = \omega \left( 1 \pm n^{-\Omega(\epsilon)} \right) \]
Our pseudo-moments are:

\[ \widetilde{E}[x_A] = \sum_{T \subseteq {[n]\atop 2}} \binom{\omega}{n} |V(T) \cup A| \chi_T(G) \]

such that \(|V(T) \cup A| \leq \tau\).
Our pseudo-moments are:

\[
\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau \\}} \left( \frac{\omega}{n} \right)^{|V(T) \cup A|} \chi_T(G)
\]

**Lemma:** If A is not a clique then

\[
\widetilde{\mathbb{E}}[x_A] = 0
\]
Our pseudo-moments are:

\[
\widehat{\mathbb{E}}[x_A] = \sum_{T \subseteq \binom{[n]}{2}} \left( \frac{\omega}{n} \right)^{|V(T) \cup A|} \chi_T(G)
\]

|V(T) \cup A| \leq \tau

**Lemma:** If A is not a clique then

\[
\widehat{\mathbb{E}}[x_A] = 0
\]

Follows from Fourier expansion of AND, and grouping terms
Our pseudo-moments are:

\[ \widetilde{\mathbb{E}}[x_A] = \sum_{T \subseteq \left[\frac{n}{2}\right]} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G) \]

with \( |V(T) \cup A| \leq \tau \)

**Lemma:** If \( A \) is not a clique then

\[ \widetilde{\mathbb{E}}[x_A] = 0 \]

Follows from Fourier expansion of AND, and grouping terms

This is why we use \( |V(T) \cup A| \leq \tau \) for truncation
Our pseudo-moments are:

\[
\tilde{E}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \ \text{with} \ |V(T) \cup A| \leq \tau}} \left( \frac{\omega}{n} \right)^{|V(T) \cup A|} \chi_T(G)
\]
Our pseudo-moments are:

\[ \tilde{\mathbb{E}}[x_A] = \sum_{T \subseteq \binom{[n]}{2}} \left( \frac{\omega}{n} \right)^{|V(T) \cup A|} \chi_T(G) \]

\[ |V(T) \cup A| \leq \tau \]

**Lemma:** Let \( f_G(x) = \sum_{|S| \leq 2d} c_A(G')x_A \) where \( \text{deg}(c_A) \leq \tau \), then

\[ \mathbb{E}[\tilde{\mathbb{E}}[f_G(x)]] = \mathbb{E}[f_G(x)] \]

\[ G \leftarrow G(n, 1/2) \quad (G, x) \leftarrow G'(n, 1/2, \omega) \]
OUTLINE

Part I: Introduction

• Planted Clique and its Applications
• The Sum-of-Squares Hierarchy
• Our Results

Part II: Fooling SOS

• The Meka-Potechin-Wigderson Moments
• Kelner’s Polynomial, and Corrections at $d = 4$
• Pseudo-Calibration and Fourier Analysis
• Symbolic Factorization and Intersection Terms
OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner’s Polynomial, and Corrections at $d = 4$
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms
What about proving positivity? e.g. $\tilde{E}[p^2] \geq 0$
What about proving positivity?  

This step is *by far* the most challenging (*as usual*)
What about proving positivity? e.g. $\mathbb{E}[p^2] \geq 0$

This step is by far the most challenging (as usual)

As is standard, it amounts to proving a certain matrix is PSD, whose entries are:

$$\mathcal{M}(I, J) = \sum_{T \subseteq \binom{[n]}{2}} \left( \frac{\omega}{n} \right)^{|V(T) \cup I \cup J|} \chi_T(G)$$

$$|V(T) \cup I \cup J| \leq \tau$$
What about proving positivity? e.g. $\tilde{\mathbb{E}}[p^2] \geq 0$

This step is by far the most challenging (as usual)

As is standard, it amounts to proving a certain matrix is PSD, whose entries are:

$$\mathcal{M}(I, J) = \sum_{\substack{T \subseteq [n] \atop |V(T) \cup I \cup J| \leq \tau}} \left( \frac{\omega}{n} \right)^{|V(T) \cup I \cup J|} \chi_T(G)$$

Goal: Write $\mathcal{M}$ as:

$$\mathcal{M} \approx \sum_{k} \mathcal{L}_k \mathcal{Q}_k \mathcal{L}_k^+$$

size of minimum vertex separator of $T$, btwn $I$ and $J$
RIBBON DECOMPOSITION

We call such graphs \((I,J)\)-Ribbons, e.g.

\[
\mathcal{R} \quad \begin{array}{c}
\text{a} \\
\text{b} \\
\text{c} \\
\text{h} \\
\text{i} \\
\text{j} \\
\text{k} \\
\text{x} \\
\text{y} \\
\text{z} \\
\end{array}
\]

with \(I = \{a, b, c\}, J = \{c, x, y, z\}\).
RIBBON DECOMPOSITION

We call such graphs \((I,J)\)-Ribbons, e.g.

with \(I = \{a, b, c\}\), \(J = \{c, x, y, z\}\). Compute leftmost and rightmost minimum vertex separators \(S_L, S_R\).
RIBBON DECOMPOSITION

We call such graphs \((I,J)\)-Ribbons, e.g.

\[
R \quad a \quad b \quad c \\
\quad b \quad h \quad i \\
\quad c \quad x \quad y \quad z
\]

with \(I = \{a, b, c\}, J = \{c, x, y, z\}\). Compute leftmost and rightmost minimum vertex separators \(S_L, S_R\). Decompose

\[
R_l \quad R_m \quad R_r \quad k \quad x \\
a \quad b \quad c \\
h \quad i \\
\quad j \\
y \quad z
\]
SYMBOLIC FACTORIZATION

Now we can write:

\[ \mathcal{M}(I, J) \approx \sum_{k} \left( \sum_{\text{valid } \mathcal{R}_l} \left( \frac{\omega}{n} \right) |V(\mathcal{R}_l)| \right) \left( \sum_{\text{valid } \mathcal{R}_m} \left( \frac{\omega}{n} \right) |V(\mathcal{R}_m)|^{-2k} \right) \left( \sum_{\text{valid } \mathcal{R}_r} \left( \frac{\omega}{n} \right) |V(\mathcal{R}_r)| \right) \]

\[ \mathcal{L}_k \quad \mathcal{Q}_k \quad \mathcal{L}_k^T \]
SYMBOLIC FACTORIZATION

Now we can write:

\[ \mathcal{M}(I, J) \approx \text{sum over } k \text{ of } \]

\[
\left( \sum_{\text{valid } \mathcal{R}_l} \left( \frac{\omega}{n} |V(\mathcal{R}_l)| \right) \right) \left( \sum_{\text{valid } \mathcal{R}_m} \left( \frac{\omega}{n} |V(\mathcal{R}_m)|^{-2k} \right) \right) \left( \sum_{\text{valid } \mathcal{R}_r} \left( \frac{\omega}{n} |V(\mathcal{R}_r)| \right) \right)
\]

\[ \mathcal{L}_k \quad \mathcal{Q}_k \quad \mathcal{L}_k^T \]

**Major issue:** \( \mathcal{R}_l, \mathcal{R}_m, \mathcal{R}_r \) were assumed to be disjoint except at \( S_L, S_R, I \cap J \) which leads to substantial error terms
SYMBOLIC FACTORIZATION

Now we can write:

\[ M(I, J) \approx \sum \text{sum over k of} \]

\[
\left( \sum_{\text{valid } R_l} \left( \frac{\omega}{n} \right) |V(R_l)| \right) \left( \sum_{\text{valid } R_m} \left( \frac{\omega}{n} \right) |V(R_m)|^{-2k} \right) \left( \sum_{\text{valid } R_r} \left( \frac{\omega}{n} \right) |V(R_r)| \right)
\]

\[ L_k \quad Q_k \quad L^T_k \]

**Major issue:** \( R_l, R_m, R_r \) were assumed to be **disjoint** except at \( S_L, S_R, I \cap J \) which leads to substantial **error terms**

**Idea:** Keep iterating the decomposition, carefully charging
ITERATING THE DECOMPOSITION

Suppose $h = j$
ITERATING THE DECOMPOSITION

Suppose $h = j$

Look for new leftmost, rightmost separators that separate $I$ from $J$ and intersection vertices
THE MAIN CHARGING ARGUMENT

Complications:

(1) Vertices can become isolated

(2) Separators not necessarily equal size

(3) Need to sum over all pre-images of ribbons, their contributions

Main Tradeoff Lemma: There is a way to tradeoff all these parameters, to charge error terms
Summary:

• Nearly optimal lower bounds against SoS, for the planted clique problem

• **Pseudo-calibration** as a recipe for constructing good pseudo-moments

• When the recipe fails, are there algorithms?

• Connections between **SoS-evidence** and **BP-evidence**?
Summary:

• Nearly optimal lower bounds against SoS, for the planted clique problem

• **Pseudo-calibration** as a recipe for constructing good pseudo-moments

• When the recipe fails, are there algorithms?

• Connections between **SoS-evidence** and **BP-evidence**?

Thanks! Any Questions?