Sum-of-Squares Lower Bounds for Hidden Clique

Yash Deshpande and Andrea Montanari

Stanford University

July 5th 2015
The Setup

$H_0$

$G \sim \mathcal{G}(n, 1/2)$
The Setup

Detection Problem

Given $G$, distinguish between $H_0$ and $H_1$.

$G \sim \mathcal{G}(n, 1/2)$

$G \sim \mathcal{G}(n, k, 1/2)$
The Setup

Detection Problem
Given $G$, distinguish between $H_0$ and $H_1$

$H_0$

$G \sim \mathbb{G}(n, 1/2)$

$H_1$

$G \sim \mathbb{G}(n, k, 1/2)$
Detection Problem
Given $G$, distinguish between $H_0$ and $H_1$
Detection Problem

Given $G$, distinguish between $H_0$ and $H_1$

$G \sim \mathbb{G}(n, 1/2)$

$G \sim \mathbb{G}(n, k, 1/2)$
The Setup

\[ A = \pm 1 \]

\[ G \sim \mathbb{G}(n, 1/2) \]

Detection Problem
Given \( A \), distinguish between \( H_0 \) and \( H_1 \)
Outline

- Motivation
Outline

- Motivation
- Spectral methods
Outline

- Motivation
- Spectral methods
- Sum of squares relaxations
Outline

- Motivation
- Spectral methods
- Sum of squares relaxations
- Main result
Outline

► Motivation

► Spectral methods

► Sum of squares relaxations

► Main result

► Overview of proof
Motivation

- Prototypical example of computational vs. statistical thresholds
Motivation

- Prototypical example of computational vs. statistical thresholds
- Reduction primitive for a variety of inference problems
Statistically…

- $k \geq C \log n$ is necessary and sufficient
Statistically...

- $k \geq C \log n$ is necessary and sufficient

- Simple second moment calculation

- [Grimmett 1975] [Arias-Castro, Verzelen 2013] [Butucea, Ingster 2011]
Spectral methods: the state of the art

\[
A = \mathbb{E}\{A\} + c1_Q1_Q^T + A - \mathbb{E}\{A\} + Z.
\]

\[
\begin{array}{cccc}
  c & c & & \\
  c & c & & \\
  & & 0 & \\
  & & 0 & 0
\end{array}
\]

\[
\mathbb{E}\{Z_{ij}\} = 0
\]

\[
\text{norm} \approx c |Q| \quad \text{norm} \approx \sqrt{n}
\]
Spectral methods: the state of the art

\[ A = \mathbb{E}\{A\} + c1_Q1_Q^T + A - \mathbb{E}\{A\} \]

\[ \mathbb{E}\{Z_{ij}\} = 0 \]

\[ \text{norm} \approx c|Q| \]

\[ |Q| = k \gtrsim \sqrt{n} \Rightarrow v_1(A) \approx 1_Q \]

[Alon et al 1998]
Sparse PCA

\[ A = \mathbb{E}\{A\} + c_1 Q 1_Q^T + A - \mathbb{E}\{A\} \]

\[ \mathbb{E}\{Z_{ij}\} = 0 \]
Sparse PCA

\[ A = \mathbb{E}\{A\} + c_1 Q_1^T + A - \mathbb{E}\{A\} \]

\[ c_1 Q_1^T \]

Sparse, Low Rank noise

\[ \mathbb{E}\{Z_{ij}\} = 0 \]
Sparse PCA

\[ A = \mathbb{E}\{A\} + c_1 q_1^T + A - \mathbb{E}\{A\} \]

\[ \mathbb{E}\{Z_{ij}\} = 0 \]

Sparse, Low Rank noise

[Berthet, Rigollet 2012] [Ma, Wu 2013] [Hajek et al 2014] [Cai et al 2015]
Summarizing...

Statistically

- \( k \geq C \log n \) sufficient
- Brute force algorithm

Computationally

- \( k \geq C \sqrt{n} \) required
- Spectral algorithm
Prior computational lower bounds

Standard reductions for NP-hardness of CLIQUE do not work
Prior computational lower bounds

Standard reductions for NP-hardness of CLIQUE do not work

- [Jerrum 1992] proved $k < n^{1/2-\varepsilon}$ hard for Metropolis
Prior computational lower bounds

Standard reductions for NP-hardness of CLIQUE do not work

- [Jerrum 1992] proved $k < n^{1/2−\varepsilon}$ hard for Metropolis

- [Feldman et al 2012]: similar result for “statistical algorithms”
Prior computational lower bounds

Standard reductions for NP-hardness of CLIQUE do not work

- [Jerrum 1992] proved $k < n^{1/2 - \varepsilon}$ hard for Metropolis

- [Feldman et al 2012]: similar result for “statistical algorithms”

- [Feige Krauthgamer 2002]: For $r$-Lovasz Schrijver:

  \[ k < c \sqrt{\frac{n}{2^r}} \text{ is hard} \]
Is $\log n \ll k \ll \sqrt{n}$ hard?

For which algorithms can we prove it?
Sum of Squares

Powerful algorithmic idea for polynomial optimization [Shor 1987], [Parrilo 2000], [Lasserre 2001]
Sum of Squares

Powerful algorithmic idea for polynomial optimization [Shor 1987], [Parrilo 2000], [Lasserre 2001]

- Principled sequence of convex relaxations
Sum of Squares

Powerful algorithmic idea for polynomial optimization
[Shor 1987], [Parrilo 2000], [Lasserre 2001]

- Principled sequence of convex relaxations
- Subsumes previous relaxation schemes (LS, SA)
Sum of Squares

Powerful algorithmic idea for polynomial optimization [Shor 1987], [Parrilo 2000], [Lasserre 2001]

▶ Principled sequence of convex relaxations

▶ Subsumes previous relaxation schemes (LS, SA)

▶ Underlies many approximation algorithms
Sum of Squares

Powerful algorithmic idea for polynomial optimization [Shor 1987], [Parrilo 2000], [Lasserre 2001]

- Principled sequence of convex relaxations
- Subsumes previous relaxation schemes (LS, SA)
- Underlies many approximation algorithms

Concrete class of algorithms to prove lower bounds
Sum of Squares for (Hidden) Clique

Polynomial optimization for Clique

\[
\text{maximize } \sum_{i=1}^{n} x_i
\]

subject to:

\[
x_i \in \{0, 1\} \quad \forall i
\]
\[
x_i x_j = 0 \quad \forall (i, j) \text{ not edge.}
\]

Key idea

▶ Optimize probability distributions over feasible set
▶ Probability distributions \( \approx \) (small number of) moments
Sum of Squares for (Hidden) Clique

Polynomial optimization for Clique

\[
\text{maximize } \sum_{i=1}^{n} x_i \\
\text{subject to:} \\
\begin{align*}
  x_i &\in \{0, 1\} & \forall i \\
  x_i x_j &= 0 & \forall (i, j) \text{ not edge.}
\end{align*}
\]

Key idea

- Optimize *probability distributions* over feasible set
Sum of Squares for (Hidden) Clique

Polynomial optimization for Clique

\[
\text{maximize } \sum_{i=1}^{n} x_i \\
\text{subject to:}
\]

\[
x_i \in \{0, 1\} \quad \forall i
\]

\[
x_i x_j = 0 \quad \forall (i, j) \text{ not edge.}
\]

Key idea

- Optimize probability distributions over feasible set
- Probability distributions \( \approx \) (small number of) moments
SOS relaxation

Some notation:

- $[n]$: set of first $n$ integers
SOS relaxation

Some notation:

- \([n]\): set of first \(n\) integers

- \(\binom{[n]}{d}\): set of subsets of size \(d\)
SOS relaxation

Some notation:

- $[n]$: set of first $n$ integers
- $\binom{[n]}{d}$: set of subsets of size $d$
- $\binom{[n]}{\leq d}$: set of subsets of size $\leq d$

Decision variable: $X : \binom{[n]}{\leq d} \rightarrow [0, 1]$
SOS relaxation

Some notation:

- \([n]\): set of first \(n\) integers
- \(\binom{[n]}{d}\): set of subsets of size \(d\)
- \(\binom{[n]}{\leq d}\): set of subsets of size \(\leq d\)

Decision variable: \(X : \binom{[n]}{\leq d} \rightarrow [0, 1]\)

Interpretation:
\[
X(\{i, j, k\}) \approx P(\{i, j, k\} \text{ induces clique}) = E\{x_i x_j x_k\}
\]
SOS($d$) for (Hidden) Clique

maximize \[ \sum_{i \in [n]} X\{\{i\}\} \]

subject to:

\[ X(\emptyset) = 1 \]
\[ X(S) \in [0, 1] \]
\[ X(S) = 0 \text{ when } S \text{ not clique} \]
\[ \text{Mom}(X) \succeq 0 \]
SOS($d$) for (Hidden) Clique

maximize \[ \sum_{i \in [n]} X(\{i\}) \]

subject to:
\[
\begin{align*}
X(\emptyset) &= 1 \\
X(S) &\in [0, 1] \\
X(S) &= 0 \text{ when } S \text{ not clique} \\
\text{Mom}(X) &\succeq 0
\end{align*}
\]

\[
\text{Mom}(X)_{S_1, S_2} = X(S_1 \cup S_2). \quad S_1, S_2 \in \binom{[n]}{\leq d/2}
\]
A Sum of Squares test

\[ \text{SOS}(G; d) = \text{optimum value of SOS}(d) \text{ program} \]

\[ T_{\text{SOS}}(G; d) = \begin{cases} 
1 & \text{if } \text{SOS}(G; d) \geq k \\
0 & \text{o.w.} 
\end{cases} \]
Main Result: hardness for SOS(4)

Theorem (Deshpande, Montanari 2015)

\[ T_{SOS}(G; 4) \text{ fails to distinguish } H_0 \text{ and } H_1 \text{ when } \]

\[ k \lesssim n^{1/3} \]

\[^1 \lesssim \text{ hides log factors}\]
Main Result: hardness for $\text{SOS}(4)$

**Theorem (Deshpande, Montanari 2015)**

$T_{\text{SOS}}(G; 4)$ fails to distinguish $H_0$ and $H_1$ when

$$k \lesssim n^{1/3}$$

Independently,

**Theorem (Meka, Potechin, Wigderson 2015)**

$T_{\text{SOS}}(G; d)$ fails to distinguish $H_0$ and $H_1$ when

$$k \lesssim n^{1/d}$$

---

$^1 \lesssim$ hides log factors
Comments

- Conjecture: $T_{SOS}(G; d)$ fails when $k \lesssim C(d)n^{1/2}$
Comments

- Conjecture: $T_{SOS}(G; d)$ fails when $k \lesssim C(d)n^{1/2}$

- Previously, [Meka, Wigderson 2013] claimed this result

- Unfortunately, proof was flawed :(
Proof strategy

Error rate = \( \mathbb{P}_0\{ T_{SOS}(G; d) = 1 \} + \mathbb{P}_1\{ T_{SOS}(G; d) = 0 \} \)
Proof strategy

Error rate = \( P_0\{ T_{SOS}(G; d) = 1 \} + P_1\{ T_{SOS}(G; d) = 0 \} \)

- By definition, \( P_1\{ T_{SOS}(G; d) = 0 \} = 0 \)

- Also \( P_0\{ T_{SOS}(G; d) = 1 \} = P_0\{ SOS(G; d) \geq k \} \)
Proof strategy

Error rate = \( \mathbb{P}_0 \{ T_{SOS}(G; d) = 1 \} + \mathbb{P}_1 \{ T_{SOS}(G; d) = 0 \} \)

- By definition, \( \mathbb{P}_1 \{ T_{SOS}(G; d) = 0 \} = 0 \)

- Also \( \mathbb{P}_0 \{ T_{SOS}(G; d) = 1 \} = \mathbb{P}_0 \{ \text{SOS}(G; d) \geq k \} \)

- Feasible point under null \( \Rightarrow \) lower bound on detection
The witness for $d = 4$

Let $G(S) = \mathbb{I}(S \text{ induces clique in } G)$

$$X(\emptyset) = 1$$
The witness for $d = 4$

Let $\mathcal{G}(S) = \mathbb{I}(S \text{ induces clique in } G)$

\[
\begin{align*}
X(\emptyset) &= 1 \\
X(S) &= \alpha(S)\mathcal{G}(S)
\end{align*}
\]
The witness for $d = 4$

Let $G(S) = \mathbb{I}(S \text{ induces clique in } G)$

$$X(\emptyset) = 1$$
$$X(S) = \alpha_{|S|} G(S)$$
The witness for $d = 4$

Let $G(S) = \mathbb{I}(S \text{ induces clique in } G)$

\[
X(\emptyset) = 1 \\
X(S) = \alpha_{|S|}G(S)
\]

- Satisfies linear constraints automatically
- Key condition: $\text{Mom}(X) \succeq 0$
The witness for $d = 4$

<table>
<thead>
<tr>
<th></th>
<th>$\emptyset$</th>
<th>1, 2</th>
<th>...</th>
<th>$n$</th>
<th>${1, 2}$</th>
<th>...</th>
<th>${n-1, n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>1</td>
<td>$\alpha_1$</td>
<td></td>
<td></td>
<td>$\alpha_2 G_{ij}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>$\alpha_2 G_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\alpha_1$</td>
<td></td>
<td></td>
<td>$\alpha_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>${1, 2}$</td>
<td>$\alpha_2 G_{ij}$</td>
<td></td>
<td>$\alpha_2 G_{ij}$ or $\alpha_3 G_{ijk}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2 G_{ij}$</td>
<td></td>
<td>$\alpha_2 G_{ij}$ or $\alpha_3 G_{ijk}$</td>
<td></td>
<td>$\alpha_3 G_{ijk}$ or $\alpha_4 G_{ijkl}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2 G_{ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\alpha_2 G_{ij}$</td>
</tr>
<tr>
<td>${n-1, n}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The witness for $d = 4$
Comparing

\[
\begin{array}{ccccc}
1, 2 & \cdots & n \\
1 & \alpha_2 G_{ij} & & & \\
\vdots & & \alpha_1 & & \\
n & & & \alpha_2 G_{ij} & \\
\end{array}
\]

\[
\begin{array}{cccc}
\{1, 2\} & \cdots & \{n-1, n\} \\
\{1, 2\} & \alpha_3 G_{ijk} \text{ or } \alpha_4 G_{ijk\ell} & & \\
\vdots & & \alpha_2 G_{ij} & \\
\{n-1, n\} & & & \\
\end{array}
\]
Comparing

$\alpha_1 G_{ij}$

$\alpha_2 G_{ij}$

$\alpha_3 G_{ijk}$ or $\alpha_4 G_{ijk\ell}$

$\alpha_2 G_{ij}$

$\alpha_3 G_{ijk}$

$\alpha_4 G_{ijk\ell}$

$\alpha_1$

$\alpha_2$

$\alpha_3$

$\alpha_4$

$\{1, 2\}$

$\{n - 1, n\}$

$\{1, 2\}$

$\{n - 1, n\}$

$\{1, 2\}$

$\{n - 1, n\}$

- Independent entries

- Highly dependent entries
Comparing

- Independent entries
- Spectrum of adj. matrix
- Highly dependent entries
- Generalized adj. matrix

\[ \alpha_1 G_{ij} \]

\[ \alpha_2 G_{ij} \]

\[ \alpha_3 G_{ijk} \text{ or } \alpha_4 G_{ijkl} \]

\[ \alpha_2 G_{ij} \]

\[ \alpha_3 G_{ijk} \text{ or } \alpha_4 G_{ijkl} \]

\[ \alpha_1 \]

\[ \alpha_2 \]

\[ \alpha_3 \]

\[ \alpha_4 \]
Comparing

- Independent entries
- Spectrum of adj. matrix
- Standard tools of RMT

- Highly dependent entries
- Generalized adj. matrix
- Not well understood
A toy example

\[ M \in \mathbb{R}^{([n]/2) \times ([n]/2)} \]

\[ M_{\{i,j\},\{k,\ell\}} = A_{ik}A_{i\ell}A_{jk}A_{j\ell} \]
A toy example

\[ M \in \mathbb{R}^{\binom{n}{2} \times \binom{n}{2}} \]

\[ M\{i,j\},\{k,\ell\} = A_{ik}A_{i\ell}A_{jk}A_{j\ell} \]

Question

How do we control \( \|M\| \)?
The Moment Method

\[ \mathbb{E}\|N\| \leq (\mathbb{E}\text{Tr}\{N^{2r}\})^{1/2r} \]

\[ N = A \quad \text{and} \quad N = M \]
The Moment Method

$$\mathbb{E}\|N\| \leq (\mathbb{E}\text{Tr}\{N^{2r}\})^{1/2r}$$

\[\begin{align*}
N &= A \\
N &= M
\end{align*}\]

\[\begin{align*}
r &= 1 \\
r &= 1
\end{align*}\]
The Moment Method

$$\mathbb{E}\|N\| \leq (\mathbb{E}\text{Tr}\{N^{2r}\})^{1/2r}$$

\[ N = A \]

\[ r = 1 \]

\[ r = 2 \]

\[ N = M \]

\[ r = 1 \]

\[ r = 2 \]
Conclusion

- Can generalize to other models
- Non-standard random matrices
- Open problem: $k \lesssim \sqrt{n}$ hard for SOS($d$)?

Thanks!
Conclusion

- Can generalize to other models
Conclusion

- Can generalize to other models
- Non-standard random matrices

\[ k \lesssim \sqrt{n} \text{ hard for } \text{SOS}(d) \]
Conclusion

- Can generalize to other models
- Non-standard random matrices
- Open problem: $k \lesssim \sqrt{n}$ hard for SOS($d$)?

Thanks!