Lower bounds on the size of semidefinite programs

James Lee  
*Univ of Washington.*

Prasad Raghavendra  
*U.C.Berkeley.*

David Steurer  
*Cornell University*
Extended formulations of polytopes

Traveling salesman problem (TSP)

given $n \times n$ cost matrix $D = (d_{ij})$, find minimum cost $n$-tour

Traveling salesman polytope

$$\text{tsp}_n = \text{convex-hull } \mathbf{1}_E(C) \in \{0,1\}^{n\times2} \quad C \text{ is } n\text{-tour}$$

**Characterization:** solving TSP on $n$ cities same as optimizing linear functions over $\text{tsp}_n$

$tsp_n$ has exponential number of facets $\implies$ no "direct" LP algorithms
Extended formulations of polytopes

**idea:** reduce optimizing linear functions over \( \text{tsp} \downarrow n \) to optimizing linear functions over polytope defined by *few linear inequalities*

**Size-\( R \) extended LP formulation for \( TSP \downarrow n \)**

*size-\( R \) polytope \( P \), defined by \( \leq R \) linear inequalities, such that \( \text{tsp} \downarrow n \) is image of \( P \) under some linear map \( \ell \)*

\[
\begin{array}{ccc}
\text{tsp} \downarrow n & \xleftarrow{\text{linear map } \ell} & P \\
\mathbb{R}^{n^2} & \text{for } R \gg n^2 & \mathbb{R}^R
\end{array}
\]

**size-\( R \) SDP algorithm for \( \text{tsp} \downarrow n \) (aka size-\( R \) extended SDP formulation)**

*size-\( R \) spectrahedron \( P \), defined by \( R \times R \) linear matrix inequality, such that \( \text{tsp} \downarrow n \) is image of \( P \) under some linear map \( \ell \)*

\[
\begin{array}{ccc}
\text{tsp} \downarrow n & \xleftarrow{\text{linear map } \ell} & P \\
\mathbb{R}^{n^2} & \text{for } R \gg n^2 & \mathbb{R}^R
\end{array}
\]
**complicated polytopes can have simple lifts**

*(here: complicated = many inequalities; simple = few inequalities)*

**unit $\ell_1$-ball**

\[
\{ \sum_{i=1}^{n} |x_i| \leq 1 \mid x \in \mathbb{R}^n \}
\]

**projection**

\[
(\text{Id} \& 0)
\]

\[
\{-y \leq x \leq y \sum_{i=1}^{n} y_i \leq 1 \mid x, y \in \mathbb{R}^n \}
\]

**comparison:** $2^n$ linear inequalities vs. $2n+1$ linear inequalities

**idea:** introduce variables for absolute values $|x|$

**other polytopes:** spanning trees, Held–Karp TSP

LP / SDP hierarchies introduce new variables systematically
lower bounds on extended LP/SDP formulations

minimum size of LP/SDP algorithms for $\text{tsp} \downarrow n$

**symmetric LP:** $2 \uparrow \Omega(n)$  
[Yannakakis]

**symmetric SDP:** $2 \uparrow \Omega(n)$  
[Lee-Raghavendra-S.-Tan Fawzi-Saunderson-Parrio]

**general LP:** $2 \uparrow \Omega(n)$  
[Fiorini-Massar-Pokutta-Tiwary-de Wolf Rothvoss]

**general SDP:** $2 \uparrow n \uparrow 1/13$  
[this talk]

Similar lower bounds for the $\text{CUT} \downarrow n$ and $\text{Correlation} \downarrow n$ polytopes.

*Unconditional lower bounds* for restricted but powerful model of computation.
GW SDP for MaxCut

Semidefinite Program for MaxCut:
[Goemans-Williamson 94]

Maximize $\frac{1}{4} \sum_{(i,j) \in E} |v_i - v_j|^2$

Subject to $|v_i|^2 = 1$

Spectrahedron:

- $Y$ is a $n \times n$-p.s.d matrix
  
  $Y_{ii} = 1$ for all $i \in [n]$

Maximize:

$\frac{1}{4} \sum_{(i,j) \in E} (Y_{ii} + Y_{jj} - 2Y_{ij})$
Generic SDP for MaxCut

Maximize $\langle w, Y \rangle$
subject to
$Y \in \text{spectrahedron } S$

$S = S \downarrow \mathbb{R}^+ \cap \{ Y | AY = b \}$

$S \downarrow \mathbb{R}^+ = \text{cone of } \mathbb{R} \times \mathbb{R} \text{ p.s.d matrices}$

Affine space $AY = b$ ($Y \in \text{size of SDP} = \mathbb{R} \uparrow \mathbb{R} \times \mathbb{R}$)
Expressing objective value:
For all integer assignments \( x \in \{-1,1\}^n \), the corresponding solution \( Q \downarrow x \in S \).

Maximize \( \langle w \downarrow G , Y \rangle \) subject to \( Y \in \text{spectrahedron } S \)
\[
S = S_{\downarrow R^+} \cap \{ Y | AY = b \}
\]

Every graph \( G \) has a ‘linearization’ \( w \downarrow G \in R^+ R \times R \)

General SDP Relaxation for MaxCut (on \( n \) vertices).
**Result (Informal Statement):**

For every Max-CSP (like MaxCut),

The $k$-round Lasserre/low-degree SOS SDP relaxation achieves the best approximation among all SDP relaxations of roughly the same size.

**Theorem:** For every integer $k \in \mathbb{N}$, for every Max-CSP,

A SDP relaxation of size $\frac{n^k}{C}$ is no more powerful than a degree $k$-SoS SDP relaxation (C = absolute constant)

Using known lower-bounds against low-degree SoS hierarchies,

**Corollary:** For every integer $k$, for every Max-CSP,

a SDP relaxation of size $\frac{n^k}{C}$ cannot yield the following approximations:

- $\frac{7}{8} + \epsilon$-approximation for Max-3-SAT for any constant $\epsilon$.
- $\frac{1}{2} + \epsilon$-approximation for Max-3-LIN for any constant $\epsilon$. 
Prior Work
(Linear Programming Extended Formulations)

- For `symmetric LPs’, an exponential lower bound for exact TSP and exact non-bipartite matching.  
  [Yannakakis 89]

- For general LPs, an exponential lower bound for exact TSP.  
  [Fiorini-etal]

- For general LPs, an exponential lower bound for exact Perfect Matching.  
  [Rothvoss]

In a slightly-more restrictive model,

- A $2^{\Omega(n^{1/2} \epsilon)}$ -lower bound for $n^{1/2} - \epsilon$ -approximation to Max-Clique.  
  [Fiorini etal]

- A $2^{\Omega(n^{1-\epsilon})}$ -lower bound for $n^{1-\epsilon}$ -approximation to Max-Clique.
Prior Work
(Linear Programming Extended Formulations)

**Theorem:** [Chan-Lee-R-Steurer]
For every integer $k < \log n / \log \log n$, for every Max-CSP, an LP relaxation of size $n^k$ is no more powerful than a $O(k)$-round Sherali-Adams Linear Program.
Low Degree Sum-of-Squares SDP Hierarchies

(Lasserre/Parrillo SDP hierarchy)
Revisiting MaxCut
Semidefinite Program

Integer Program:
Domain: $x_{1}, x_{2}, x_{3}, ..., x_{n} \in \{-1,1\}$
(x for vertex i)

Maximize:
$\frac{1}{4} \sum_{\{i,j\} \in \mathcal{E}} (x_{i} - x_{j})^2$
(Number of Edges Cut)

Convex Extension of Integer Program:

Domain: Probability density $D$ over assignments $x \in \{-1,1\}^{n}$

Maximize:
$\mathbb{E}[D(x) \cdot \frac{1}{4} \sum_{\{i,j\} \in \mathcal{E}} (x_{i} - x_{j})^2 ]$
(Expected Number of Edges Cut under $D$)
Degree d Sum-of-Squares SDP

Convex Extension of Integer Program:

Domain: Probability density $D$ over assignments $x \in \{-1,1\}^n$

Maximize: $E \downarrow x \left[D(x) \cdot \frac{1}{4} \sum_{(i,j) \in E \uparrow \downarrow} (x \downarrow i - x \downarrow j)^2 \right]$  
(Expected Number of Edges Cut under $D$)

Subject to: $E \downarrow x D(x) = 1$

$D(x) \geq 0 \quad \forall x \in \{0,1\}^n$

$E \downarrow x D(x) \cdot p \uparrow 2 \quad (x) \geq 0$

$\forall p \text{ with } \deg(p) \leq d$

$D = \text{``pseudodensity''}$
Sum of Squares Proofs

**Degree d Sum-of-Squares SDP**

Maximize:

\[ E \downarrow x [D(x) \cdot \frac{1}{4} \sum (i,j) \in E \uparrow \updownarrow (x \downarrow i - x \downarrow j)^2 ] \]

Subject to:

\[ E \downarrow x D(x) = 1 \]

\[ E \downarrow x [D(x) \cdot p \uparrow 2 (x)] \geq 0 \]

\[ \forall p \text{ with } \deg(p) \leq d \]

**Dual SDP**

Minimize \( c \) such that,

\[ c - G(x) = \sum j \uparrow \downarrow p \downarrow j \uparrow 2 (x) \]

for some degree d polynomials \( p \downarrow j \uparrow 2 \)

**Definition (SoS degree):**

Given \( f: \{0,1\}^n \to R \),

\[ \deg_{sos} (f) = \text{minimum } d \text{ such that } f = \sum j \uparrow \downarrow p \uparrow 2 (x) \]

for some \( \{ p \downarrow j \} \) with \( \deg(p \downarrow j \uparrow 2) \leq d \)
Degree $d$ Sum-of-Squares SDP

Maximize:
\[
E\downarrow x \left[ D(x) \cdot \frac{1}{4} \sum_{(i,j) \in E\uparrow \downarrow} (x\downarrow_i - x\downarrow_j)^2 \right]
\]

Subject to:
\[
E\downarrow x \ D(x) = 1
\]
\[
E\downarrow x \left[ D(x) \cdot p\uparrow \downarrow^2 (x) \right] \geq 0
\]
for some degree $d$ polynomials $p\downarrow j\uparrow \downarrow$

Dual SDP

Minimize $c$

such that,
\[
c\!-\!G(x)\!=\!\sum_{j\uparrow \downarrow} p\downarrow j\uparrow \downarrow^2 (x)
\]

for some degree $d$ polynomials $p\downarrow j\uparrow \downarrow$

Definition (SoS degree):
Given $f:\{0,1\}^n \to \mathbb{R}$,
\[
sosdeg(f) = \text{minimum } d \text{ such that } f = \sum_{j\uparrow \downarrow} p\downarrow j\uparrow \downarrow^2 (x)
\]
for some $\{p\downarrow j\}$ with $\deg(p\downarrow j\uparrow \downarrow) \leq d$
Duality

Degree d Sum-of-Squares SDP
Maximize:
\[ E \downarrow x \ [D(x) \cdot \frac{1}{4} \sum (i,j) \in E \uparrow \downarrow (x_i - x_j)^2 ] \]
Subject to:
\[ E \downarrow x \ D(x) = 1 \]
\[ E \downarrow x [D(x) \cdot p \uparrow \downarrow (x)] \geq 0 \]
\[ \forall p \text{ with } \deg(p) \leq d \]

Dual SDP
Minimize \[ c \]
such that,
\[ \text{sosdeg}(c - G(x)) = d \]

By strong duality,
\[ SDPOPT(G) > c \iff \text{sosdeg}(c - G(x)) > d \]
\[ \text{cone}(\{ p \downarrow \uparrow \downarrow | \deg(p \downarrow \uparrow \downarrow ) \leq d \}) \]
Yannakakis’ characterization of extension complexity
**Theorem:** For any graph $G$, $\text{SDP-OPT}$ if and only if

Maximize $(w \downarrow G, Y)$
subject to $Y \in \text{spectrahedron } S$

$S = S \downarrow R^+ \cap \{Y | AY = b\}$
**SoS characterization of SDPs**

Maximize \( (w \downarrow G, Y) \)

subject to

\( Y \in \text{spectrahedron } S \)

\[ S = S \downarrow R \uparrow + \cap \{ Y | AY = b \} \]

For \( d \)-round Lasserre SDP,

\[ V = \{ \text{vector space of polynomials of degree } \leq d \} \]

**Vector Space of functions (\( \{-1,1\}^n \rightarrow \mathbb{R} \))**

\[ V = \text{span}\{ v \downarrow \uparrow 1, \ldots, v \downarrow R \uparrow R \} \]

Minimize \( c \)

such that

\[ c - G(x) = \sum q \in V \downarrow \uparrow \deg q \downarrow 2 \ (x) \]

**Degree d SOS SDP for MaxCut**

Minimize \( c \)

such that

\[ c - \sum (i,j) \in E \downarrow \uparrow w \downarrow ij (x \downarrow i - x \downarrow j) \uparrow 2 = \sum \deg (q) \leq d \downarrow \uparrow q(x) \uparrow 2 \quad \text{for } x \in \{-1,1\}^n \]
**PSD rank**

**Definition:**
PSD rank(M) = smallest $R$ so that

\[ \exists \text{ factorization: } M = P \cdot Q^\top \]

where rows of $P$, $Q$ are $R \times R$ p.s.d matrices

**Theorem:** [Yannakakis]

\[ \exists \text{ size-}R \text{ SDP with approx. ratio } \alpha \text{ for } n \text{-variable MAXCUT} \]

\[ \iff \text{psd-rank}(M) \leq R \text{ for } \]

variable assignments $x \in \{0,1\}^n$

**MAXCUT instances $G$ in $n$ variables**

\[ M(G,x) = \alpha \cdot \max(G) - G(x) \]
Theorem: [Yannakakis]

\[ \exists \text{ size-} R \text{ SDP with approx. ratio } \alpha \text{ for } n\text{-variable MAXCUT} \]
\[ \iff \text{ psd-rank}(M) \leq R \text{ for } \]
\[ \text{variable assignments } x \in \{0,1\}^n \]

MAXCUT instances \( G \) in \( n \) variables

\[ M(G,x) = \alpha \cdot \max(G) - G(x) \]

Proof: Suppose following SDP gives an \( \alpha \)-approximation.

Maximize \( \langle w \downarrow G, Y \rangle \)
subject to
\( Y \in \text{Spectrahedron } S \)
\( S = S \downarrow R \uparrow + \cap \{ Y | AY = b \} \)

For each assignment \( x \in \{0,1\}^n \rightarrow Q \downarrow x \in S \downarrow R \uparrow + \)

For all graphs \( G \)
\[ G(x) = \langle w \downarrow G, Q \downarrow x \rangle \]
**nonneg.-rank and psd-rank**

factorization: \( M = P \cdot Q^\top \)

two ways to evaluate:

1. inner products of rows of \( P \) and \( Q \)
2. outer products of columns of \( P \) and \( Q \)

*for nonneg.-rank:*

nonneg. rows \( \iff \) nonneg. columns

\( \rightarrow \) can use both ways to evaluate factorization

all known lower bounds work with outer-product view (nonnegative rectangles)

*for psd-rank*

psd rows \( \not\iff \) psd columns
Main Technical Result
**main theorem**

**main theorem**: suppose $f$ has sos deg > $d$. then exists $C \downarrow f \geq 1$ such that

$$\text{psd-rank}(M \downarrow n \uparrow f) \geq C \downarrow f \cdot n \uparrow d / 5$$

for all $n \in \mathbb{N}$

points $x \in \{0,1\}^n$

subsets $S \subseteq [n]$ with $|S| = m$

evaluations of $f$ applied to subcubes

$M \downarrow n \uparrow f (S, x) = f(x \downarrow S)$

$x \downarrow S = (x \downarrow s_1, \ldots, x \downarrow s_m)$ for $S = \{s_1, \ldots, s_m\}$
**Main theorem:** suppose $f$ has sos deg $> d$. then exists $C \downarrow f \geq 1$ such that $\text{psd-rank}(M \downarrow n \uparrow f) \geq C \downarrow f \cdot n \uparrow d / 5$ for all $n \in \mathbb{N}$

---

**Main theorem $\Rightarrow$ Optimality of Lasserre SDP**

**Proof:** Suppose degree $d$ Lasserre has an $\alpha$-integrality gap for MaxCut

$\Leftrightarrow \exists$ some graph $G$ such that $\deg \downarrow \text{sos} (\alpha \cdot \max(G) - G(x)) > d$

$\Leftrightarrow$ For $f = \alpha \cdot \max(G) - G(x)$

$\text{psd-rank}(M \downarrow n \uparrow f) \geq C \downarrow f \cdot n \uparrow d / 5$

$\Leftrightarrow$ No SDP of size $o(n \uparrow d / 5)$ gets an $\alpha$–approximation.
main theorem: suppose \( f \) has sos deg > \( d \). then exists \( C \downarrow f \geq 1 \) such that
\[
\text{psd-rank}(M \downarrow n \uparrow f) \geq C \downarrow f \cdot n \uparrow d / 5 \quad \text{for all } n \in \mathbb{N}
\]

subsets \( S \subseteq [n] \) with \( |S| = m \)

points \( x \in \{0,1\}^n \)

evaluations of \( f \) applied to subcubes

\[
M \downarrow n \uparrow f (S,x) = f(x \downarrow S)
\]

\( x \downarrow S = (x \downarrow s_1, \ldots, x \downarrow s_m) \) for \( S = \{s_1, \ldots, s_m\} \)

sanity check: \( M \downarrow n \uparrow f \) does not have low-deg. factorizations \((\text{Tr } P \downarrow S Q \downarrow x) \downarrow S, x\)

suppose \( Q \downarrow x = R(x) \uparrow 2 \) and \( x \mapsto R(x) \) has degree at most \( d/2 \)

then, \( \text{Tr } P \downarrow S Q \downarrow x = \|P \downarrow S \uparrow 1/2 \cdot R(x)\| \downarrow F \uparrow 2 \)

\( \rightarrow \) sum of squares of degree-\( d/2 \) polynomials for each \( S \)

\( \rightarrow \) cannot be factorization of \( M \downarrow n \uparrow f \) (because \( f \) has sos deg > \( d \))
Separating from low-degree factorizations

By duality, \( \text{sos-deg}(f) > d \) if and only if

\[ \exists \text{ deg.-}d \text{ pseudo-distr. } D \text{ with } \mathbb{E} \downarrow x D(x)f(x) < 0 \]

\[
\text{cone}(\{p \downarrow j \uparrow 2 \mid \text{deg}(p \downarrow j \uparrow 2) \leq d\})
\]

Idea: Use the linear functional \( \langle D, N \rangle = E \downarrow S, x D(x \downarrow S)N(S, x) \) to separate \( M \) from all low-psd rank matrices?

Know: \[ \mathbb{E} \downarrow S, x D(x \downarrow S)M \downarrow n \uparrow f(S, x) < -\varepsilon \]

To show: \[ \mathbb{E} \downarrow S, x D(x \downarrow S) \cdot \text{Tr } P \downarrow S Q \downarrow x \geq -\varepsilon \text{ for all candidate factor. } (\text{Tr } P \downarrow S Q \downarrow x) \downarrow S, x \]
main theorem: suppose $f$ has sos deg $> d$. then exists $C \downarrow f \geq 1$ such that
\[ \text{psd-rank}(M \downarrow n \uparrow f) \geq C \downarrow f \cdot n \uparrow d / 5 \] for all $n \in \mathbb{N}$

points $x \in \{0,1\} \uparrow n$
evaluations of $f$

subsets $S \subseteq [n]$ with $|S| = m$
evaluations of $f$ applied to subcubes

strategy: approximate general low-rank factorization by low-deg. factor.

1. bootstrapping: general low-rank factor. $\rightarrow$ moderate deg. factor. (deg. $n \uparrow \varepsilon$)

2. boosting: $M \downarrow n \uparrow f$ does not have moderate deg. factorization

idea: quantum entropy maximization / learning; quite mysterious

idea: random restriction; quite standard
**step 2: boosting via random restriction**

let $D:\{0,1\}^m\to\mathbb{R}$ be deg.-$d$ pseudo-distribution

let $(\text{Tr } P\downarrow S Q\downarrow x)\downarrow S, x$ be any rank-$n\uparrow d/10$ psd factorization with max / avg $= O(1)$

suppose $Q\downarrow x = R(x)\uparrow 2$ and $x\mapsto R(x)$ has deg $\leq n\uparrow \varepsilon$.

then, $\mathbb{E}\downarrow S, x D(x\downarrow S) \cdot \text{Tr } P\downarrow S Q\downarrow x \geq -o\downarrow D, m$ (1)

**sanity check:** conclusion true if $R(x)$ has deg $\leq d/2$

**observation:** enough if $x\mapsto P\downarrow S \uparrow 1/2 R(x)$ has deg $\leq d/2$ in $S$

**idea:** deg-$n\uparrow \varepsilon$ polynomial very close to deg-$d/2$ after restriction to random set $S$

$$\mathbb{P}\downarrow S \left(\text{deg}−n\uparrow \varepsilon \text{ character } \chi\downarrow T \text{ has deg}>d/2 \text{ in } S \right)$$

$$= \mathbb{P}\downarrow S \left( |T\cap S|>d/2 \right) \leq (|S| \cdot |T|/n)\uparrow d/2 \approx n\uparrow−d/2$$

small enough to counter effect of $P\downarrow S$
Thank You