Semialgebraic Proofs and Efficient Algorithm Design

Noah Fleming

Department of Computer Science

University of Toronto

Joint work with Pravesh Kothari and Toni Pitassi

Sum-of-Squares

Sum-of-Squares: Powerful proof system

— Proofs correspond to a family of SDPs

Sum-of-Squares

Sum-of-Squares: Powerful proof system
— Proofs correspond to a family of SDPs

Sum-of-Squares has become a popular tool in algorithm design

Sum-of-Squares

Sum-of-Squares: Powerful proof system
— Proofs correspond to a family of SDPs

Sum-of-Squares has become a popular tool in algorithm design

Powerful:
- Captures many famous approximation algorithms for NP hard problems such as the Goemans Williamson algorithm for MaxCut
- Gives optimal approximations of any CSP under the Unique Games Conjecture [Raghavendra08]

Sum-of-Squares

Sum-of-Squares: Powerful proof system
— Proofs correspond to a family of SDPs

Sum-of-Squares has become a popular tool in algorithm design

Powerful:
- Captures many famous approximation algorithms for NP hard problems such as the Goemans Williamson algorithm for MaxCut
- Gives optimal approximations of any CSP under the Unique Games Conjecture [Rag08]

Simple Algorithm Design Strategy:
- Sum-of-Squares proofs are automatizable.
- Proofs that a solution exist automatically give efficient algorithms for finding that solution. Main difficulty is rounding the solution.

Outline

1. Developing the Sum-of-Squares Relaxation
2. Phrasing the Relaxation as an SDP
3. The Dual Sum-of-Squares Proofs and Completeness
4. Convergence and Strong Duality
5. Upper Bounds
6. Lower Bounds

Outline

1. Developing the Sum-of-Squares Relaxation
2. Phrasing the Relaxation as an SDP
3. The Dual Sum-of-Squares Proofs and Completeness
4. Convergence and Strong Duality
5. Upper Bounds
6. Lower Bounds

Polynomial Optimization Problems

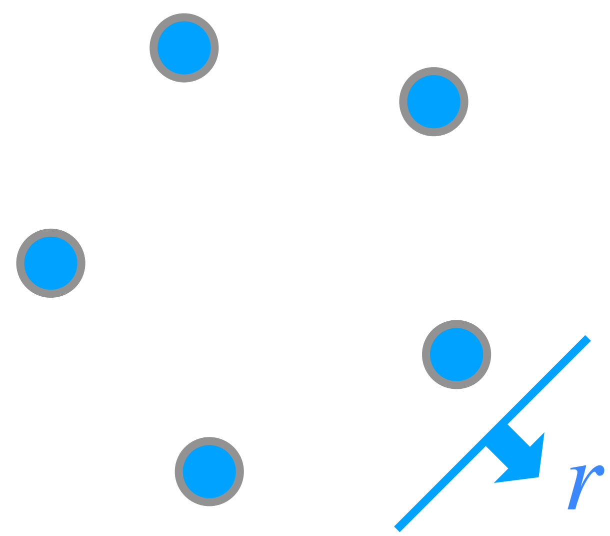
$\mathscr{P} \subseteq \mathbb{R}[x]$ a set of polynomials, $r \in \mathbb{R}[x]$ linear.

$$\begin{aligned} \max_x \quad & r(x) \\ \text{s.t.} \quad & p(x) \geq 0 \qquad \forall p \in \mathscr{P} \end{aligned}$$

Polynomial Optimization Problems

$\mathscr{P} \subseteq \mathbb{R}[x]$ a set of polynomials, $r \in \mathbb{R}[x]$ linear.

$$\max_x \quad r(x)$$
$$\text{s.t.} \quad p(x) \geq 0 \qquad \forall p \in \mathscr{P}$$

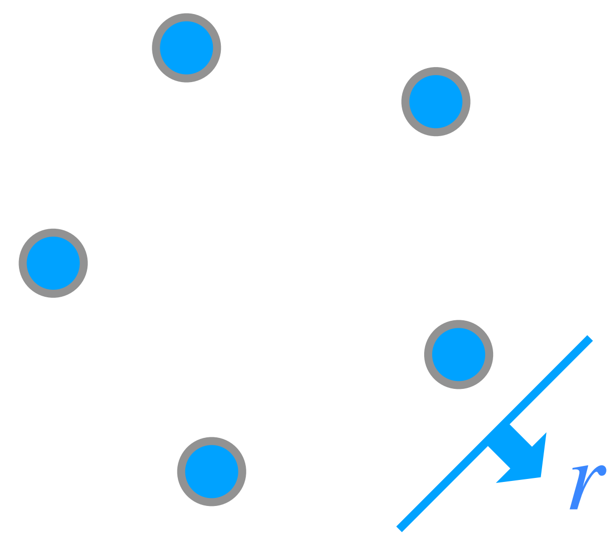


$$K_{\mathscr{P}} := \{x \in \mathbb{R}^n : p_i(x) \geq 0 \, \forall p_i \in \mathscr{P}\}$$

Polynomial Optimization Problems

$\mathscr{P} \subseteq \mathbb{R}[x]$ a set of polynomials, $r \in \mathbb{R}[x]$ linear.

$$\begin{aligned} \max_x \quad & r(x) \\ \text{s.t.} \quad & p(x) \geq 0 \qquad \forall p \in \mathscr{P} \end{aligned}$$



$$K_{\mathscr{P}} := \{x \in \mathbb{R}^n : p_i(x) \geq 0 \, \forall p_i \in \mathscr{P}\}$$

Problem: Polynomial optimization problems are NP-hard to solve in general.

Goal: Develop a tractable relaxation that achieves good approximations to many problems we care about

Motivating the SoS Relaxation

Goal: Develop a tractable relaxation that achieves good approximations to many problems we care about.

Standard approach is via convex programming.

Motivating the SoS Relaxation

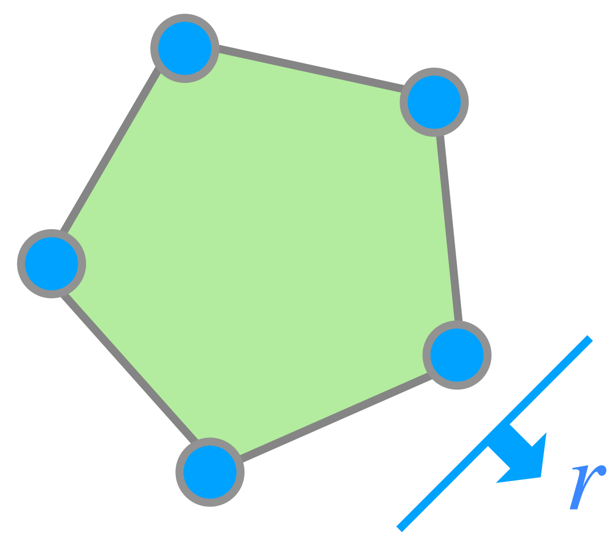
Goal: Develop a tractable relaxation that achieves good approximations to many problems we care about.

Standard approach is via convex programming.

Thought Experiment:

Take the convex relaxation of $K_{\mathscr{P}}$

$conv(K_{\mathscr{P}})$



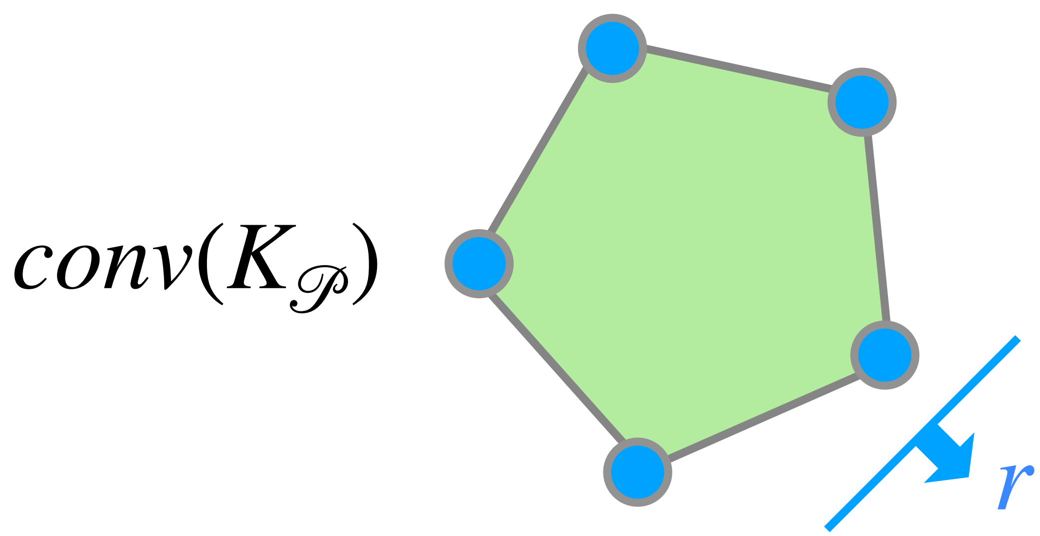
Motivating the SoS Relaxation

Goal: Develop a tractable relaxation that achieves good approximations to many problems we care about.

Standard approach is via convex programming.

Thought Experiment:

Take the convex relaxation of $K_{\mathscr{P}}$



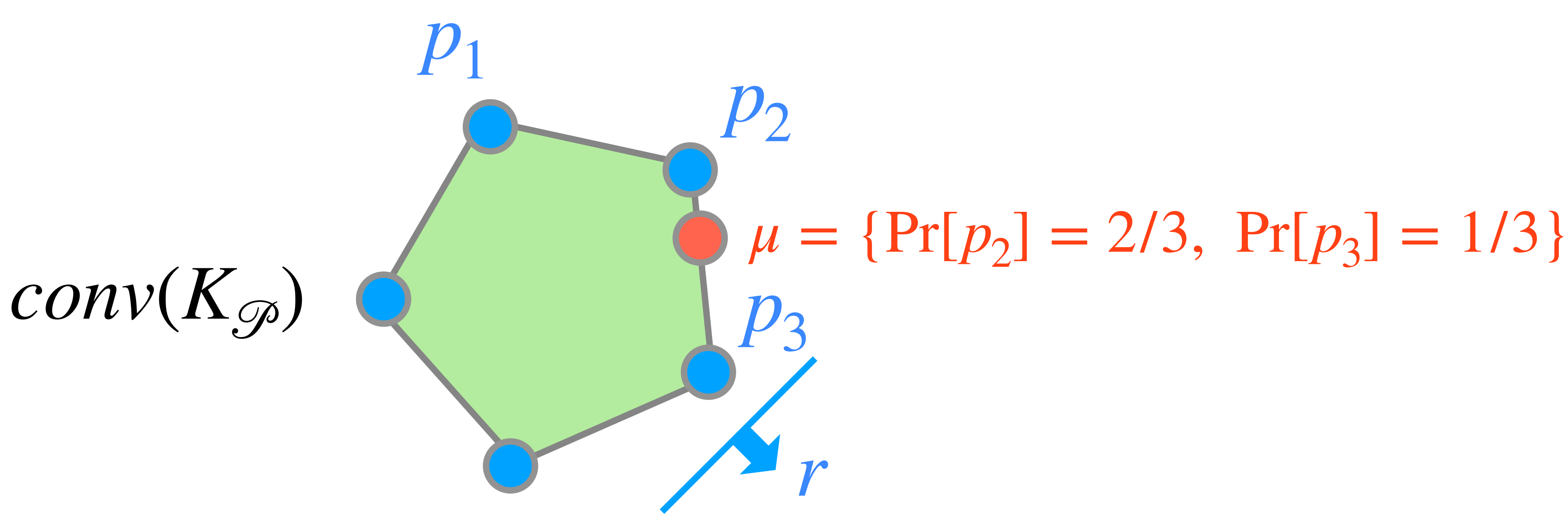
$$\max_{x \in K_{\mathscr{P}}} r(x) = \max_{x \in conv(K_{\mathscr{P}})} r(x)$$

By linearity of $r(x)$, any optimal solution $x \in conv(\mathscr{P})$ is a convex combination of optimal $x \in K_{\mathscr{P}}$

Motivating the SoS Relaxation

Goal: Develop a tractable relaxation that achieves good approximations to many problems we care about.

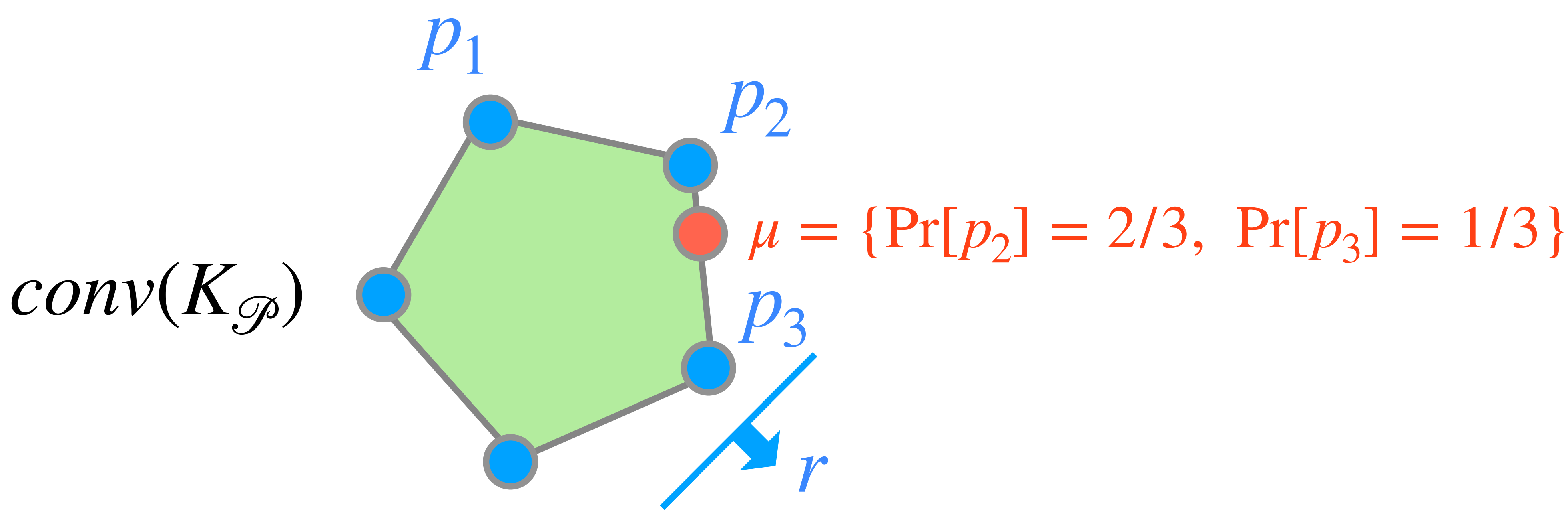
Distributional View: view the points in $conv(K_{\mathscr{P}})$ as distributions μ supported on the points $K_{\mathscr{P}}$



Motivating the SoS Relaxation

Goal: Develop a tractable relaxation that achieves good approximations to many problems we care about.

Distributional View: view the points in $conv(K_{\mathscr{P}})$ as distributions μ supported on the points $K_{\mathscr{P}}$



$$\max_{x \in K_{\mathscr{P}}} r(x) = \max_{x \in conv(K_{\mathscr{P}})} r(x) = \max \mathbb{E}_{\mu}[r(x)] : \mu \textbf{ is supported on } K_{\mathscr{P}}$$

Motivating the SoS Relaxation

Distributions μ can be described by their moments $\mathbb{E}_\mu[x^I]$
where $x^I := \Pi_{i \in I} x_i$

Motivating the SoS Relaxation

Distributions μ can be described by their moments $\mathbb{E}_\mu[x^I]$ where $x^I := \Pi_{i \in I} x_i$

Suggests a relaxation

Relaxation: restrict attention to the degree $\leq d$ moments of these distributions, $\mathbb{E}[x^I]$ for $|I| \leq d$

– Only n^d such moments

Motivating the SoS Relaxation

Distributions μ can be described by their moments $\mathbb{E}_\mu[x^I]$ where $x^I := \Pi_{i \in I} x_i$

Suggests a relaxation

Relaxation: restrict attention to the degree $\leq d$ moments of these distributions, $\mathbb{E}[x^I]$ for $|I| \leq d$

– Only n^d such moments

However...

NP-hard to determine if there exists a distribution μ on $K_{\mathscr{P}}$ which agrees with a given set of moments $\{\mathbb{E}[x^I]\}_{|I| \leq d}$

Motivating the SoS Relaxation

Distributions μ can be described by their moments $\mathbb{E}_\mu[x^I]$ where $x^I := \Pi_{i \in I} x_i$

Suggests a relaxation

Relaxation: restrict attention to the degree $\leq d$ moments of these distributions, $\mathbb{E}[x^I]$ for $|I| \leq d$

— Only n^d such moments

However...

NP-hard to determine if there exists a distribution μ on $K_{\mathscr{P}}$ which agrees with a given set of moments $\{\mathbb{E}[x^I]\}_{|I| \leq d}$

Therefore

Look for efficient tests which distinguish collections of moments which belong to distributions supported on K_P

The Sum-of-Squares Relaxation

$\{\mathbb{E}[x^I]\}_{|I|\leq d}$ = linear function $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$.

The Sum-of-Squares Relaxation

$\{\mathbb{E}[x^I]\}_{|I|\leq d}$ = linear function $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$.

Want:
A set of efficient tests distinguishing $\tilde{\mathbb{E}}$ that agree with the moments of a true distribution on $K_{\mathscr{P}}$ from those that do not.

The Sum-of-Squares Relaxation

$\{\mathbb{E}[x^I]\}_{|I|\leq d}$ = linear function $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$.

Want:
A set of efficient tests distinguishing $\tilde{\mathbb{E}}$ that agree with the moments of a true distribution on $K_{\mathscr{P}}$ from those that do not.

Obvious tests of consistency:

- $\tilde{\mathbb{E}}[q^2(x)] \geq 0 \qquad \forall q \in \mathbb{R}[x]_{\leq d/2}$

The Sum-of-Squares Relaxation

$\{\mathbb{E}[x^I]\}_{|I|\leq d}$ = linear function $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$.

Want:
A set of efficient tests distinguishing $\tilde{\mathbb{E}}$ that agree with the moments of a true distribution on $K_{\mathscr{P}}$ from those that do not.

Obvious tests of consistency:

- $\tilde{\mathbb{E}}[q^2(x)] \geq 0 \qquad \forall q \in \mathbb{R}[x]_{\leq d/2}$
- $\tilde{\mathbb{E}}[p(x)] \geq 0 \qquad \forall p \in \mathscr{P}$

The Sum-of-Squares Relaxation

$\{\mathbb{E}[x^I]\}_{|I|\leq d}$ = linear function $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$.

Want:
A set of efficient tests distinguishing $\tilde{\mathbb{E}}$ that agree with the moments of a true distribution on $K_{\mathscr{P}}$ from those that do not.

Obvious tests of consistency:

- $\tilde{\mathbb{E}}[q^2(x)] \geq 0 \qquad \forall q \in \mathbb{R}[x]_{\leq d/2}$
- $\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0 \qquad \forall p \in \mathscr{P}, \forall q \in \mathbb{R}[x]_{\leq (d-deg(p))/2}$

The Sum-of-Squares Relaxation

$\{\mathbb{E}[x^I]\}_{|I|\le d}$ = linear function $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\le d} \to \mathbb{R}$.

Want:
A set of efficient tests distinguishing $\tilde{\mathbb{E}}$ that agree with the moments of a true distribution on $K_{\mathscr{P}}$ from those that do not.

Obvious tests of consistency:

- $\tilde{\mathbb{E}}[q^2(x)] \ge 0 \qquad \forall q \in \mathbb{R}[x]_{\le d/2}$
- $\tilde{\mathbb{E}}[p(x)q^2(x)] \ge 0 \qquad \forall p \in \mathscr{P}, \forall q \in \mathbb{R}[x]_{\le (d-deg(p))/2}$

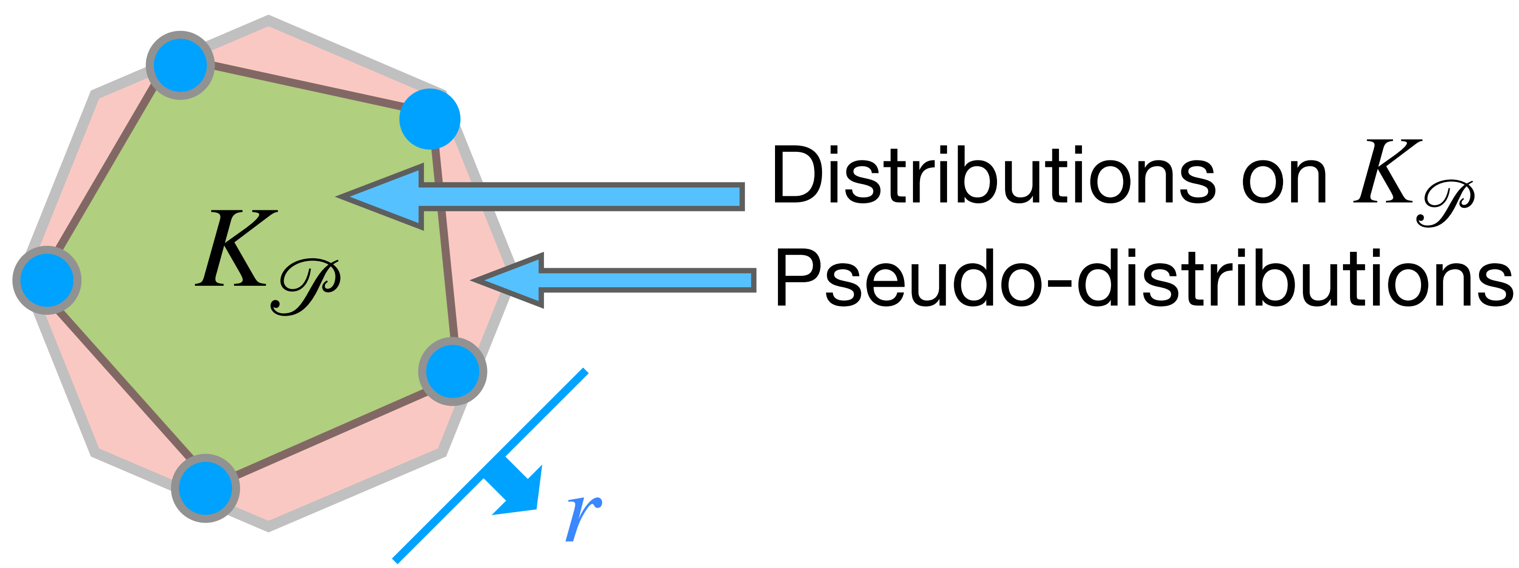
Degree-d Pseudo-Expectation for $\mathscr{P}$: Any linear function $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\le d} \to \mathbb{R}$ satisfying

1. $\tilde{\mathbb{E}}[1] = 1$
2. $\tilde{\mathbb{E}}[q^2(x)] \ge 0 \qquad \forall q \in \mathbb{R}[x]_{\le d/2}$
3. $\tilde{\mathbb{E}}[p(x)q^2(x)] \ge 0 \quad \forall q \in \mathbb{R}[x]_{\le (d-deg(p)/2}, p \in \mathscr{P}$

The Sum-of-Squares Relaxation

Degree-d Pseudo-Expectation for $\mathscr{P}$: Any linear function $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$ satisfying

1. $\tilde{\mathbb{E}}[1] = 1$
2. $\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$
3. $\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq (d-deg(p)/2}, \; p \in \mathscr{P}$



The Sum-of-Squares Relaxation

Degree-d Pseudo-Expectation for $\mathscr{P}$: Any linear function $\tilde{\mathbb{E}} : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$ satisfying

1. $\tilde{\mathbb{E}}[1] = 1$
2. $\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$
3. $\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq (d-deg(p)/2}, \ p \in \mathscr{P}$

The Sum-of-Squares Relaxation

$\max \tilde{\mathbb{E}}[r(x)]$

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

$\tilde{\mathbb{E}}[q^2(x)] \geq 0$ for all $q \in \mathbb{R}[x]_{\leq d/2}$

$\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0$ for all $p \in \mathscr{P}$, $q \in \mathbb{R}[x]_{\leq (d-deg(p))/2}$

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

n^d variables, one for each monomial.

Outline

1. Developing the Sum-of-Squares Relaxation
2. Phrasing the Relaxation as an SDP
3. The Dual Sum-of-Squares Proofs and Completeness
4. Convergence and Strong Duality
5. Upper Bounds
6. Lower Bounds

Solving the Relaxation

$\max \tilde{\mathbb{E}}[r(x)]$

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

$\tilde{\mathbb{E}}[q^2(x)] \geq 0$ for all $q \in \mathbb{R}[x]_{\leq d/2}$

$\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0$ for all $p \in \mathscr{P}$, $q \in \mathbb{R}[x]_{\leq (d - deg(p))/2}$

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

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

Solving the Relaxation

$\max \tilde{\mathbb{E}}[r(x)]$
s.t. $\tilde{\mathbb{E}}[1] = 1$
$\tilde{\mathbb{E}}[q^2(x)] \geq 0$ for all $q \in \mathbb{R}[x]_{\leq d/2}$
$\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0$ for all $p \in \mathscr{P}$, $q \in \mathbb{R}[x]_{\leq (d-deg(p))/2}$
$\tilde{\mathbb{E}}$ is linear

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

Idea: rewrite polynomials as vector products
—Square polynomials become PSD constraints.

Solving the Relaxation

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

Monomial vector: v_d where $(v_d)_I = x^I$ for $|I| \leq d$

Any $p \in \mathbb{R}[x]_{\leq d}$ can be written as

$$p(x) = \overrightarrow{p}^T v_d(x)$$

$\overrightarrow{p}$ is the coefficient vector of the monomials in $p(x)$

For $n = 2$:

$$x_1^2 + 3x_2 + 4 = \overset{\overrightarrow{p}}{\begin{bmatrix} 4 & 0 & 3 & 1 & 0 & 0 \end{bmatrix}} \overset{\overrightarrow{v_2}}{\begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix}}$$

Solving the Relaxation

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

Monomial vector: v_d where $(v_d)_I = x^I$ for $|I| \leq d$

Any $p \in \mathbb{R}[x]_{\leq d}$ can be written as

$$p(x) = \overrightarrow{p}^T v_d(x)$$

$\overrightarrow{p}$ is the coefficient vector of the monomials in $p(x)$

Rephrase $\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$:

$$\tilde{\mathbb{E}}[q^2(x)] = \tilde{\mathbb{E}}[\overrightarrow{q} v_d^T v_d \overrightarrow{q}^T] = \overrightarrow{q} \tilde{\mathbb{E}}[v_d^T v_d] \overrightarrow{q}^T \geq 0$$

Solving the Relaxation

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

Monomial vector: v_d where $(v_d)_I = x^I$ for $|I| \leq d$

Any $p \in \mathbb{R}[x]_{\leq d}$ can be written as

$$p(x) = \overrightarrow{p}^T v_d(x)$$

$\overrightarrow{p}$ is the coefficient vector of the monomials in $p(x)$

Rephrase $\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$:

$\tilde{\mathbb{E}}[q^2(x)] = \tilde{\mathbb{E}}[\overrightarrow{q} v_d^T v_d \overrightarrow{q}^T] = \overrightarrow{q}\,\tilde{\mathbb{E}}[v_d^T v_d]\,\overrightarrow{q}^T \geq 0$ PSD constraint!

Solving the Relaxation

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

Monomial vector: v_d where $(v_d)_I = x^I$ for $|I| \leq d$

Any $p \in \mathbb{R}[x]_{\leq d}$ can be written as

$$p(x) = \overrightarrow{p}^T v_d(x)$$

$\overrightarrow{p}$ is the coefficient vector of the monomials in $p(x)$

Rephrase $\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$:

$\tilde{\mathbb{E}}[q^2(x)] = \tilde{\mathbb{E}}[\overrightarrow{q} v_d^T v_d \overrightarrow{q}^T] = \overrightarrow{q} \tilde{\mathbb{E}}[v_d^T v_d] \overrightarrow{q}^T \geq 0$ PSD constraint!

Moment Matrix: $(M_d)_{|I|,|J| \leq d/2} = \tilde{\mathbb{E}}[x^{I+J}]$, then $M_d = \tilde{\mathbb{E}}[v_d^T v_d]$

Solving the Relaxation

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

Monomial vector: v_d where $(v_d)_I = x^I$ for $|I| \leq d$

Any $p \in \mathbb{R}[x]_{\leq d}$ can be written as

$$p(x) = \overrightarrow{p}^T v_d(x)$$

$\overrightarrow{p}$ is the coefficient vector of the monomials in $p(x)$

Rephrase $\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$:

$\tilde{\mathbb{E}}[q^2(x)] = \tilde{\mathbb{E}}[\overrightarrow{q} v_d^T v_d \overrightarrow{q}^T] = \overrightarrow{q}\tilde{\mathbb{E}}[v_d^T v_d]\overrightarrow{q}^T \geq 0$ PSD constraint!

Moment Matrix: $(M_d)_{|I|,|J| \leq d/2} = \tilde{\mathbb{E}}[x^{I+J}]$, then $M_d = \tilde{\mathbb{E}}[v_d^T v_d]$

$\tilde{\mathbb{E}}[q^2(x)] \geq 0 \ \ \forall q \in \mathbb{R}[x]_{\leq d/2}$ becomes $M_d \succeq 0$

Solving the Relaxation

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

$\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$ becomes $M_d \succeq 0$

$$M_2 = \begin{bmatrix} \tilde{\mathbb{E}}[1], & \tilde{\mathbb{E}}[x_1], & \ldots, & \tilde{\mathbb{E}}[x_n] \\ \tilde{\mathbb{E}}[x_1], & \tilde{\mathbb{E}}[x_1 x_1], & \ldots, & \tilde{\mathbb{E}}[x_1 x_n] \\ \vdots & \vdots & \cdots & \vdots \\ \tilde{\mathbb{E}}[x_n], & \tilde{\mathbb{E}}[x_n x_1], & \ldots, & \tilde{\mathbb{E}}[x_n x_n] \end{bmatrix}$$

Solving the Relaxation

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

$\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$ becomes $M_d \succeq 0$

Rephrase $\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0 \quad \forall p \in \mathscr{P}, q \in \mathbb{R}[x]_{\leq (d - deg(p))/2}$:

Solving the Relaxation

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

$\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$ becomes $M_d \succeq 0$

Rephrase $\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0 \quad \forall p \in \mathscr{P}, q \in \mathbb{R}[x]_{\leq (d-deg(p))/2}$:

Moment Matrix for $p \in \mathscr{P}$:

$M_d^p := \tilde{\mathbb{E}}[p(x)v_{d'}v_{d'}^T]$ where $(M_d^p)_{I,J} = \sum_{|K| \leq deg(p)} p_K \tilde{\mathbb{E}}[x^{I+J+K}]$

$|I|, |J| \leq d'$ where $d' = (d - deg(p))/2$

$\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq (d-deg(p))/2}$ becomes $M_d^p \succeq 0$

Solving the Relaxation

Goal: Phrase as an SDP of size $|\mathscr{P}| \cdot n^{O(d)}$

$\tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq d/2}$ becomes $M_d \succeq 0$

Rephrase $\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0 \quad \forall p \in \mathscr{P}, q \in \mathbb{R}[x]_{\leq (d-deg(p))/2}$:

Moment Matrix for $p \in \mathscr{P}$:

$M_d^p := \tilde{\mathbb{E}}[p(x)v_{d'}v_{d'}^T]$ where $(M_d^p)_{I,J} = \sum_{|K| \leq deg(p)} p_K \tilde{\mathbb{E}}[x^{I+J+K}]$

$|I|, |J| \leq d'$ where $d' = (d - deg(p))/2$

$\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq (d-deg(p))/2}$ becomes $M_d^p \succeq 0$

SoS SDP Relaxation

$$SOS_d(\mathscr{P}) \left\{ \begin{array}{lll} \max & \tilde{\mathbb{E}}[r(x)] & \\ \text{s.t.} & M_d \succeq 0 & \\ & M_d^p \succeq 0 & \forall p \in \mathscr{P} \\ & \tilde{\mathbb{E}}[1] = 1 & \end{array} \right\}$$

$|\mathscr{P}| \cdot n^{O(d)}$ size SDP

Solving the Relaxation

$$SOS_d(\mathscr{P}) \left\{ \begin{array}{ll} \max & \tilde{\mathbb{E}}[p(x)] \\ \text{s.t.} & M_d \succeq 0 \\ & M_d^p \succeq 0 \qquad \forall p \in \mathscr{P} \\ & \tilde{\mathbb{E}}[1] = 1 \end{array} \right.$$

Solvable by the Ellipsoid Method in time $|\mathscr{P}|\, n^{O(d)} \log(1/\varepsilon)$ to within an additive error ε

Solving the Relaxation

$$SOS_d(\mathscr{P}) \begin{cases} \max & \tilde{\mathbb{E}}[p(x)] \\ \text{s.t.} & M_d \succeq 0 \\ & M_d^p \succeq 0 \quad \forall p \in \mathscr{P} \\ & \tilde{\mathbb{E}}[1] = 1 \end{cases}$$

Solvable by the Ellipsoid Method in time $|\mathscr{P}|n^{O(d)}\log(1/\varepsilon)$ to within an additive error ε

A solution to $SOS_d(\mathscr{P})$ is on n^d variables.

Obtain an approximate solution to $\mathscr{P}$ by projecting to $[n]$

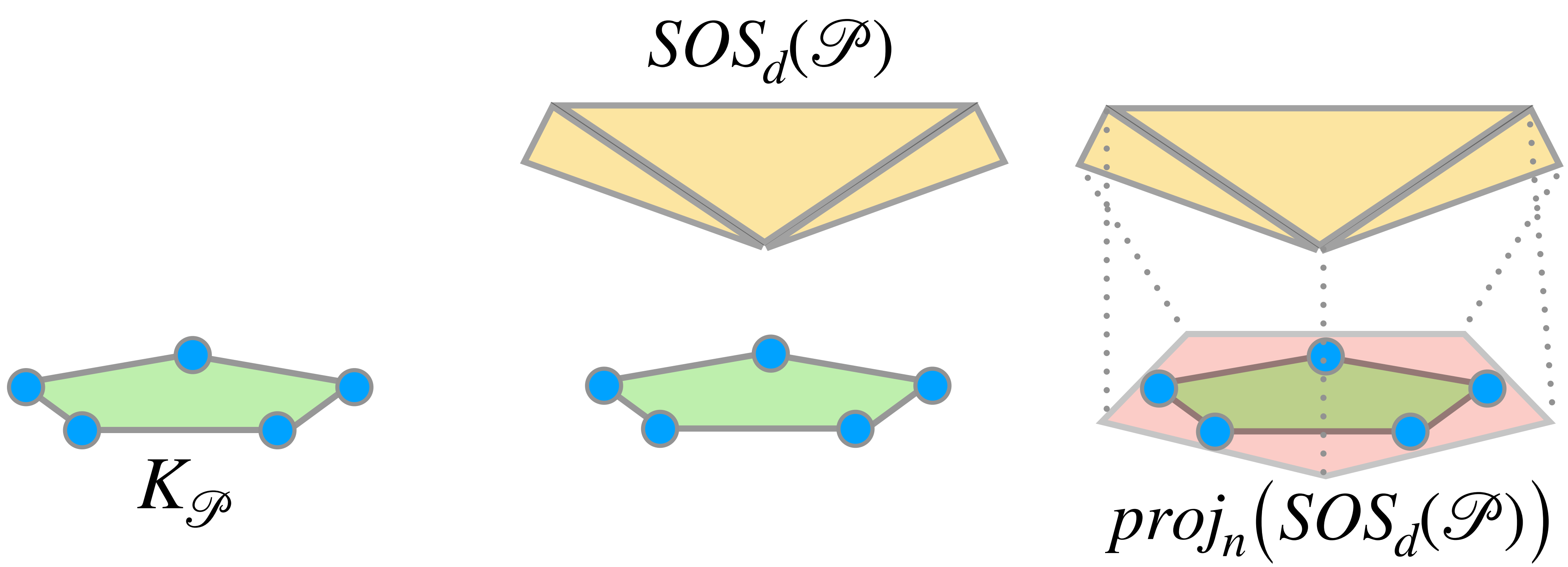
Solving the Relaxation

$$SOS_d(\mathscr{P}) \begin{cases} \max & \tilde{\mathbb{E}}[p(x)] \\ \text{s.t.} & M_d \succeq 0 \\ & M_d^p \succeq 0 \quad \forall p \in \mathscr{P} \\ & \tilde{\mathbb{E}}[1] = 1 \end{cases}$$

Solvable by the Ellipsoid Method in time $|\mathscr{P}|n^{O(d)}\log(1/\varepsilon)$ to within an additive error ε

A solution to $SOS_d(\mathscr{P})$ is on n^d variables.

Obtain an approximate solution to $\mathscr{P}$ by projecting to $[n]$



Max Cut

Max Cut POP

$$\max \sum_{i<j} w_{i,j}(x_i - x_j)^2$$

$$\text{s.t. } x_i^2 - x_i \geq 0$$

$$x_i - x_i^2 \geq 0$$

Degree-2 SOS Relaxation

SDP Formulation

Moment Matrices

Max Cut

Max Cut POP

$$\max \sum_{i<j} w_{i,j}(x_i - x_j)^2$$

$$\text{s.t. } x_i^2 - x_i \geq 0$$

$$x_i - x_i^2 \geq 0$$

Degree-2 SOS Relaxation

$$\max \sum_{i<j} w_{i,j}\tilde{\mathbb{E}}[(x_i - x_j)^2]$$

$$\text{s.t. } \tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq 1}$$

$$\tilde{\mathbb{E}}[x_i^2 - x_i] \geq 0$$

$$\tilde{\mathbb{E}}[x_i - x_i^2] \geq 0$$

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

SDP Formulation

Moment Matrices

Max Cut

Max Cut POP

$$\max \sum_{i<j} w_{i,j}(x_i - x_j)^2$$
$$\text{s.t. } x_i^2 - x_i \geq 0$$
$$x_i - x_i^2 \geq 0$$

Degree-2 SOS Relaxation

$$\max \sum_{i<j} w_{i,j}\tilde{\mathbb{E}}[(x_i - x_j)^2]$$
$$\text{s.t. } \tilde{\mathbb{E}}[q^2(x)] \geq 0 \quad \forall q \in \mathbb{R}[x]_{\leq 1}$$
$$\tilde{\mathbb{E}}[x_i^2 - x_i] \geq 0$$
$$\tilde{\mathbb{E}}[x_i - x_i^2] \geq 0$$
$$\tilde{\mathbb{E}}[1] = 1$$

SDP Formulation

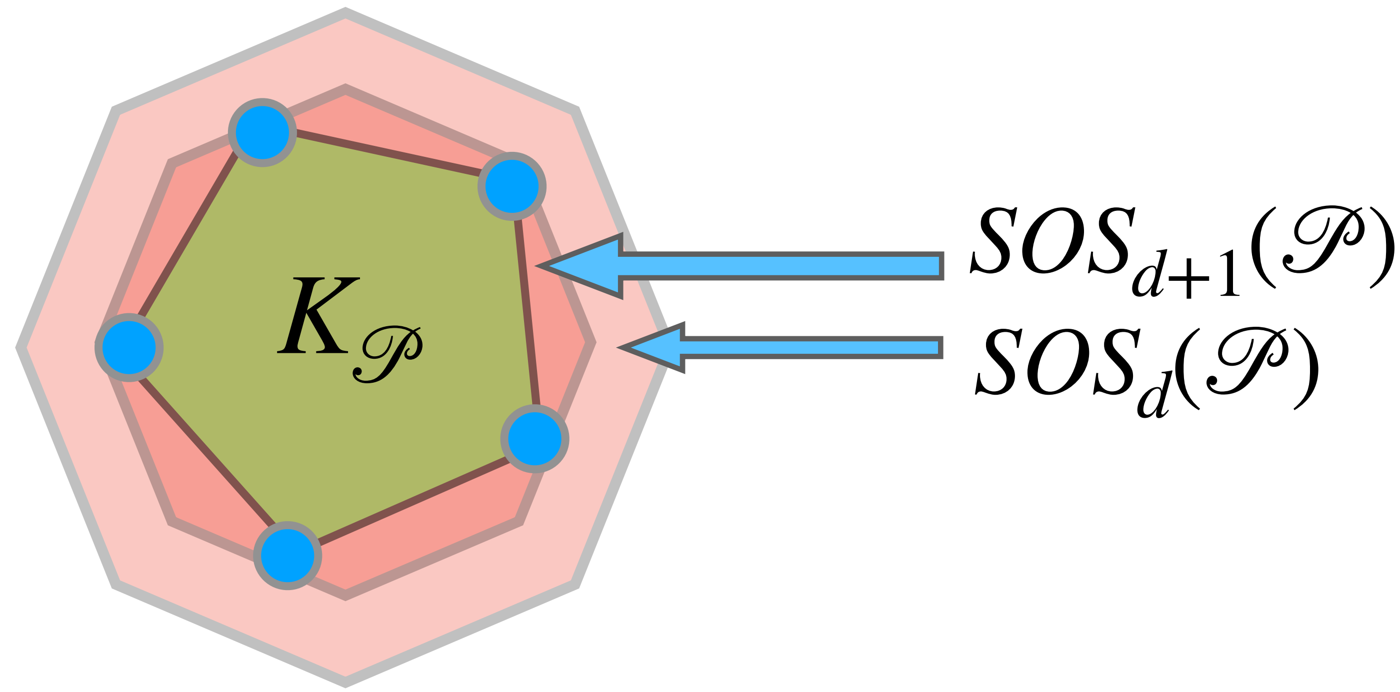
$$\max \sum_{i<j} w_{i,j}\tilde{\mathbb{E}}[(x_i - x_j)^2]$$
$$\text{s.t. } M_2 \succeq 0$$
$$M_2^{x_i - x_i^2 \geq 0} \succeq 0$$
$$M_2^{x_i^2 - x_i \geq 0} \succeq 0$$
$$\tilde{\mathbb{E}}[1] = 1$$

Moment Matrices

$$M_2 = \begin{bmatrix} \tilde{\mathbb{E}}[1], & \tilde{\mathbb{E}}[x_1], & \ldots, & \tilde{\mathbb{E}}[x_n] \\ \tilde{\mathbb{E}}[x_1], & \tilde{\mathbb{E}}[x_1x_1], & \ldots, & \tilde{\mathbb{E}}[x_1x_n] \\ \vdots & \vdots & \ldots & \vdots \\ \tilde{\mathbb{E}}[x_n], & \tilde{\mathbb{E}}[x_nx_1], & \ldots, & \tilde{\mathbb{E}}[x_nx_n] \end{bmatrix}$$
$$M_2^{x_i^2 - x_i} = \tilde{\mathbb{E}}[x_i^2 - x_i]$$
$$M_2^{x_i - x_i^2} = \tilde{\mathbb{E}}[x_i - x_i^2]$$

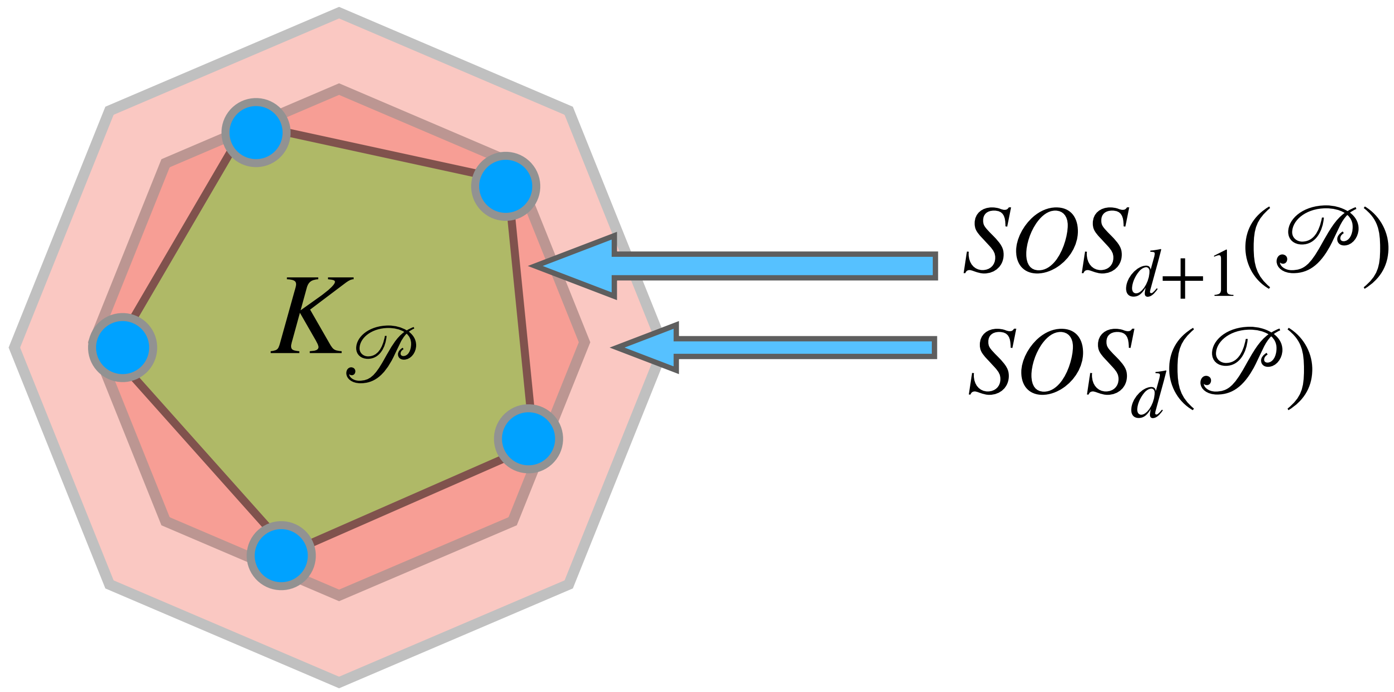
Hierarchy of Relaxations

The Sum-of-Squares relaxations form a hierarchy of ever-tightening spectahedrons parameterized by the degree d of the relaxation



Hierarchy of Relaxations

The Sum-of-Squares relaxations form a hierarchy of ever-tightening spectahedrons parameterized by the degree d of the relaxation



Can we guarantee convergence to $K_{\mathscr{P}}$?

—Not known to be true in General.

—We will see later that convergence can be guaranteed under certain assumptions on $\mathscr{P}$. This follows from duality.

Outline

1. Developing the Sum-of-Squares Relaxation
2. Phrasing the Relaxation as an SDP
3. The Dual Sum-of-Squares Proofs and Completeness
4. Convergence and Strong Duality
5. Upper Bounds
6. Lower Bounds

Certifying a Good Solution

Given an SoS relaxation, how can we certify an upper bound on its object?

Certifying a Good Solution

Given an SoS relaxation, how can we certify an upper bound on its object? Duality!

—Find the minimum $\lambda \in \mathbb{R}$ such that $\lambda - r(x)$ is non-negative over $SOS_d(\mathscr{P})$

Certifying a Good Solution

Given an SoS relaxation, how can we certify an upper bound on its object? Duality!

—Find the minimum $\lambda \in \mathbb{R}$ such that $\lambda - r(x)$ is non-negative over $SOS_d(\mathscr{P})$

Dual program corresponds to finding a good sum-of-squares decomposition of $\lambda - r(x)$

Dual:

$$\begin{aligned} &\min \lambda \\ &\text{s.t.} \quad \lambda - r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x) \\ &\qquad\quad q_p \in \mathbb{R}[x]_{\leq (d - deg(p))/2} \\ &\qquad\quad \lambda \in \mathbb{R} \end{aligned}$$

Weak Duality

$$\max \tilde{\mathbb{E}}[r(x)]$$

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

$\tilde{\mathbb{E}}[q^2(x)] \geq 0$

$\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0$

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

Primal

$$\min \lambda$$

s.t. $\lambda - r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$

$q_p \in \mathbb{R}[x]_{\leq (d - deg(p))/2}$

$\lambda \in \mathbb{R}$

Dual

Weak Duality

$$\max \tilde{\mathbb{E}}[r(x)]$$

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

$\tilde{\mathbb{E}}[q^2(x)] \geq 0$

$\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0$

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

$$\min \lambda$$

s.t. $\lambda - r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$

$q_p \in \mathbb{R}[x]_{\leq (d - deg(p))/2}$

$\lambda \in \mathbb{R}$

Weak Duality: Let $\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})$ and $r(x) = \lambda - \Sigma_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$ then $\tilde{\mathbb{E}}[r(x)] \leq \lambda$.

Weak Duality

$$\max \tilde{\mathbb{E}}[r(x)]$$
s.t. $\tilde{\mathbb{E}}[1] = 1$
$\tilde{\mathbb{E}}[q^2(x)] \geq 0$
$\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0$
$\tilde{\mathbb{E}}$ is linear

$$\min \lambda$$
s.t. $\lambda - r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x)q_p^2(x)$
$q_p \in \mathbb{R}[x]_{\leq (d - deg(p))/2}$
$\lambda \in \mathbb{R}$

Weak Duality: Let $\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})$ and $r(x) = \lambda - \Sigma_{p \in \mathscr{P} \cup \{1\}} p(x)q_p^2(x)$ then $\tilde{\mathbb{E}}[r(x)] \leq \lambda$.

Proof: $\tilde{\mathbb{E}}[r(x)] = \tilde{\mathbb{E}}[\lambda] - \sum_{p \in \mathscr{P} \cup \{1\}} \tilde{\mathbb{E}}[p(x)q_p^2(x)]$ (Linearity)

$= \lambda - \sum_{p \in \mathscr{P} \cup \{1\}} \tilde{\mathbb{E}}[p(x)q_p^2(x)]$ ($\tilde{\mathbb{E}}[1] = 1$)

$\leq \lambda$ ($\tilde{\mathbb{E}}[p(x)q_p^2(x)] \geq 0$)

Weak Duality

$$\max \tilde{\mathbb{E}}[r(x)]$$
$$\text{s.t. } \tilde{\mathbb{E}}[1] = 1$$
$$\tilde{\mathbb{E}}[q^2(x)] \geq 0$$
$$\tilde{\mathbb{E}}[p(x)q^2(x)] \geq 0$$
$\tilde{\mathbb{E}}$ is linear

$$\min \lambda$$
$$\text{s.t. } \lambda - r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x)q_p^2(x)$$
$$q_p \in \mathbb{R}[x]_{\leq (d - deg(p))/2}$$
$$\lambda \in \mathbb{R}$$

Weak Duality: Let $\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})$ and $r(x) = \lambda - \Sigma_{p \in \mathscr{P} \cup \{1\}} p(x)q_p^2(x)$ then $\tilde{\mathbb{E}}[r(x)] \leq \lambda$.

Proof: $\tilde{\mathbb{E}}[r(x)] = \tilde{\mathbb{E}}[\lambda] - \sum_{p \in \mathscr{P} \cup \{1\}} \tilde{\mathbb{E}}[p(x)q_p^2(x)]$ (Linearity)

$$= \lambda - \sum_{p \in \mathscr{P} \cup \{1\}} \tilde{\mathbb{E}}[p(x)q_p^2(x)] \qquad (\tilde{\mathbb{E}}[1] = 1)$$

$$\leq \lambda \qquad (\tilde{\mathbb{E}}[p(x)q_p^2(x)] \geq 0)$$

Writing $\lambda - r(x)$ as a degree-d sum of squares is a Sum-of-Squares proof that the maximum over $SOS_d(\mathscr{P})$ is at most λ

Sum-of-Squares Proofs

Sum-of-Squares Proof: A degree-d SoS proof of $r \in \mathbb{R}[x]$ from $\mathscr{P} \subseteq \mathbb{R}[x]$ is a set of polynomials $q_p \in \mathbb{R}[x]_{(d-deg(p))/2}$ such that

$$r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$$

Size: minimum number of bits needed to represent the proof

Sum-of-Squares Proofs

Sum-of-Squares Proof: A degree-d SoS proof of $r \in \mathbb{R}[x]$ from $\mathscr{P} \subseteq \mathbb{R}[x]$ is a set of polynomials $q_p \in \mathbb{R}[x]_{(d-deg(p))/2}$ such that

$$r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$$

Size: minimum number of bits needed to represent the proof

Sum-of-Squares Refutation: An SoS proof of -1 from $\mathscr{P}$.

- certifies that $K_{\mathscr{P}} = \emptyset$.

Sum-of-Squares Proofs

Sum-of-Squares Proof: A degree-d SoS proof of $r \in \mathbb{R}[x]$ from $\mathscr{P} \subseteq \mathbb{R}[x]$ is a set of polynomials $q_p \in \mathbb{R}[x]_{(d-deg(p))/2}$ such that

$$r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$$

Size: minimum number of bits needed to represent the proof

Sum-of-Squares Refutation: An SoS proof of -1 from $\mathscr{P}$.

- certifies that $K_{\mathscr{P}} = \emptyset$.

Weak Duality: If there exists a degree-d pseudo-expectation for $\mathscr{P}$, then there does not exist a degree-d refutation of $\mathscr{P}$.

Sum-of-Squares Proofs

Sum-of-Squares Proof: A degree-d SoS proof of $r \in \mathbb{R}[x]$ from $\mathscr{P} \subseteq \mathbb{R}[x]$ is a set of polynomials $q_p \in \mathbb{R}[x]_{(d-deg(p))/2}$ such that

$$r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$$

Size: minimum number of bits needed to represent the proof

Sum-of-Squares Refutation: An SoS proof of -1 from $\mathscr{P}$.

- certifies that $K_{\mathscr{P}} = \emptyset$.

Weak Duality: If there exists a degree-d pseudo-expectation for $\mathscr{P}$, then there does not exist a degree-d refutation of $\mathscr{P}$.

Proof: Let $-1 = \Sigma_{\mathscr{P} \cup \{1\}} p(x) q_p^2(x)$ be a degree-d refutation and $\tilde{\mathbb{E}}$ be a degree-d pseudo-expectation for $\mathscr{P}$ then

$$-1 = -\tilde{\mathbb{E}}[1] = \tilde{\mathbb{E}}[-1] = \Sigma_{p \in \mathscr{P} \cup \{1\}} \tilde{\mathbb{E}}[p(x) q_p^2(x)] \geq 0$$

Sum-of-Squares Proofs

Proofs of CNF formulas: $x_1 \vee x_2 \vee \neg x_3$ becomes $x_1 + x_2 + (1 - x_3) - 1 \geq 0$.

Also include boolean axioms $x_i^2 - x_i = 0$.

Sum-of-Squares Proofs

Proofs of CNF formulas: $x_1 \vee x_2 \vee \neg x_3$ becomes $x_1 + x_2 + (1 - x_3) - 1 \geq 0$.

Also include boolean axioms $x_i^2 - x_i = 0$.

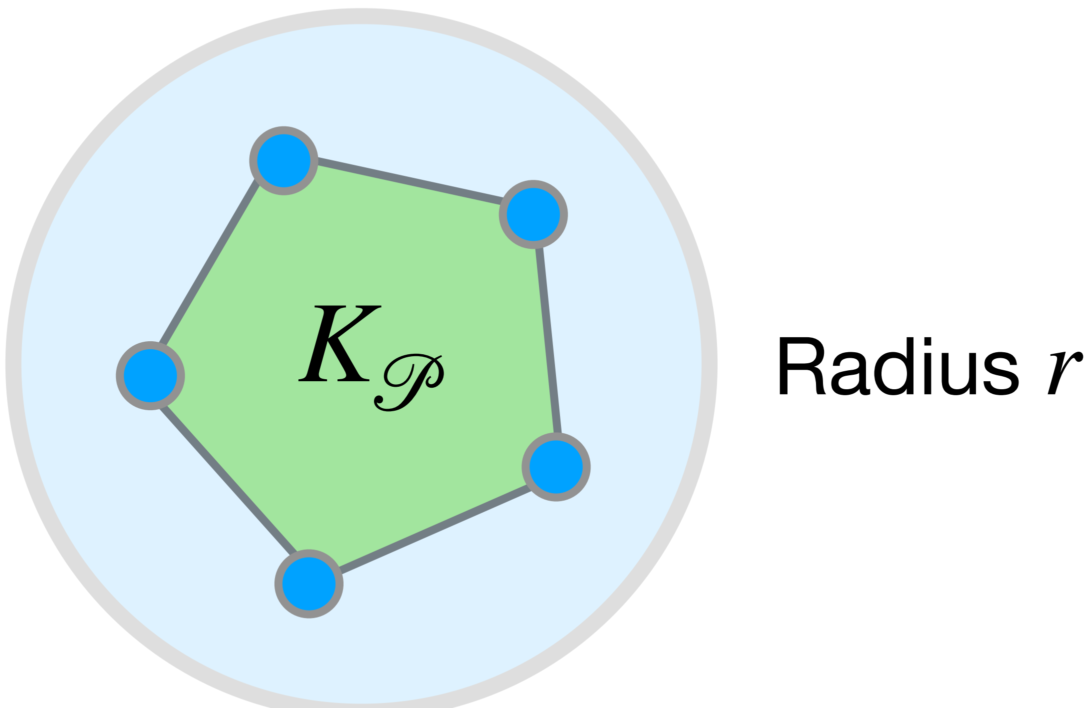
SoS is a sound and complete proof system for any set of polynomials $\mathscr{P}$ containing the boolean axioms

Sum-of-Squares Proofs

Proofs of CNF formulas: $x_1 \vee x_2 \vee \neg x_3$ becomes $x_1 + x_2 + (1 - x_3) - 1 \geq 0$.
Also include boolean axioms $x_i^2 - x_i = 0$.

SoS is a sound and complete proof system for any set of polynomials $\mathscr{P}$ satisfying the Archimedean Assumption

Archimedean Assumption: $\mathscr{P}$ contains a constraint of the form $r^2 - \Sigma_{i\in[n]} x_i^2 \geq 0$ for some r.



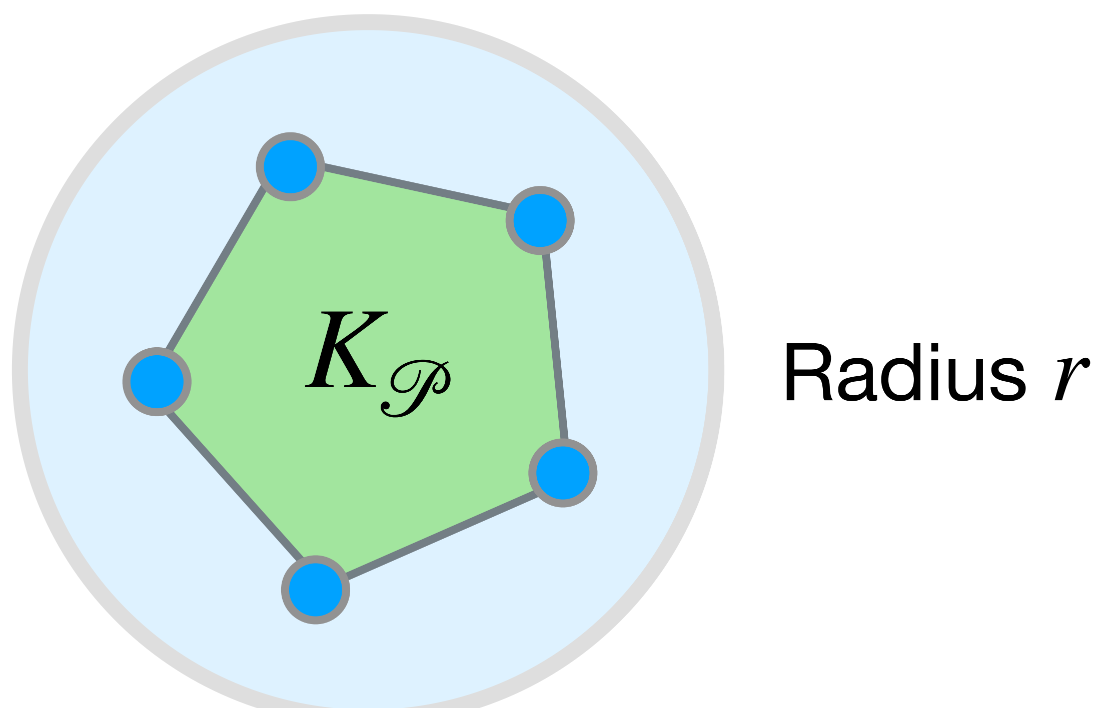
Sum-of-Squares Proofs

Proofs of CNF formulas: $x_1 \vee x_2 \vee \neg x_3$ becomes $x_1 + x_2 + (1 - x_3) - 1 \geq 0$.
Also include boolean axioms $x_i^2 - x_i = 0$.

SoS is a sound and complete proof system for any set of polynomials $\mathscr{P}$ satisfying the Archimedean Assumption

Archimedean Assumption: $\mathscr{P}$ contains a constraint of the form $r^2 - \Sigma_{i \in [n]} x_i^2 \geq 0$ for some r.

—Axioms $x_i^2 - x_i = 0$ already satisfy Archimedean Assumption



Sum-of-Squares Proofs

Proofs of CNF formulas: $x_1 \vee x_2 \vee \neg x_3$ becomes $x_1 + x_2 + (1 - x_3) - 1 \geq 0$.
Also include boolean axioms $x_i^2 - x_i = 0$.

SoS is a sound and complete proof system for any set of polynomials $\mathscr{P}$ satisfying the Archimedean Assumption

Archimedean Assumption: $\mathscr{P}$ contains a constraint of the form $r^2 - \Sigma_{i \in [n]} x_i^2 \geq 0$ for some r.

—Axioms $x_i^2 - x_i = 0$ already satisfy Archimedean Assumption

Putinar's Positivstellensatz: Let $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfy the Archimedean assumption. Then $r(x) > 0$ for all $x \in K_{\mathscr{P}}$ iff

$$r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$$

for some $q_p \in \mathbb{R}[x]$.

Sum-of-Squares Proof of PHP

Pigeonhole Principle:

a. $\Sigma_{j\in[n]} p_{i,j} - 1 \geq 0 \qquad \forall i \in [n+1]$

b. $1 - p_{i,j} - p_{i',j} \geq 0 \qquad \forall i \neq i' \in [n+1], \forall j \in [n]$

c. $p_{i,j}^2 - p_{i,j} = 0 \qquad \forall i \in [n+1], j \in [n]$

SoS Refutation of PhP:

1. Derive $1 - \Sigma_{i\in[n+1]} p_{i,j} \quad \forall j$ "Each hole has one pigeon"
2. Sum the constraints in 1 over $j \in [n]$
$$\Sigma_{j\in[n]}(1 - \Sigma_{i\in[n+1]} p_{i,j}) = n - \Sigma_{i,j} p_{i,j}$$
3. Sum the constraints in a. over $i \in [n+1]$ to get.
$$\Sigma_{i\in[n+1]}\left(\Sigma_{j\in[n]} p_{i,j} - 1\right) = \Sigma_{i,j} p_{i,j} - (n+1)$$
4. Add 2 and 3 to derive -1.

Proof of 1 as an SoS polynomial:

$$\Sigma_{i\neq i'\in[n]}(1 - p_{i,j} - p_{i',j})p_{i,j} + (1 - \Sigma_{i\in[n]} p_{i,j})^2 = 1 - \Sigma_{i\in[n]} p_{i,j}$$

Outline

1. Developing the Sum-of-Squares Relaxation
2. Phrasing the Relaxation as an SDP
3. The Dual Sum-of-Squares Proofs and Completeness
4. Convergence and Strong Duality
5. Upper Bounds
6. Lower Bounds

Convergence of the SoS hierarchy

Can we guarantee that our hierarchy of SDP relaxations converges to $K_{\mathscr{P}}$?

– Does $\lim_{d\to\infty} \max_{\tilde{E}\in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[r(x)] = \max_{x\in K_{\mathscr{P}}} r(x)$?

Convergence of the SoS hierarchy

Can we guarantee that our hierarchy of SDP relaxations converges to $K_{\mathscr{P}}$?

– Does $\lim_{d\to\infty} \max_{\tilde{E}\in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[r(x)] = \max_{x\in K_{\mathscr{P}}} r(x)$?

Convergence is holds under the Archimedean Assumption.

Archimedean Assumption: $\mathscr{P}$ contains a constraint of the form $r^2 - \Sigma_{i\in[n]} x_i^2 \geq 0$ for some r.

Convergence: Let $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfy the Archimedean Assumption

$$\lim_{d\to\infty} \max_{\tilde{E}\in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[r(x)] = \max_{x\in K_{\mathscr{P}}} r(x)$$

Convergence of the SoS hierarchy

Can we guarantee that our hierarchy of SDP relaxations converges to $K_{\mathscr{P}}$?

– Does $\lim_{d\to\infty} \max_{\tilde{E}\in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[r(x)] = \max_{x\in K_{\mathscr{P}}} r(x)$?

Convergence is holds under the Archimedean Assumption.

Archimedean Assumption: $\mathscr{P}$ contains a constraint of the form $r^2 - \Sigma_{i\in[n]} x_i^2 \geq 0$ for some r.

Convergence: Let $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfy the Archimedean Assumption

$$\lim_{d\to\infty} \max_{\tilde{E}\in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[r(x)] = \max_{x\in K_{\mathscr{P}}} r(x)$$

Proof: Combine strong duality with completeness

Convergence of the SoS hierarchy

Convergence: Let $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfy the Archimedean Assumption

$$\lim_{d \to \infty} \max_{\tilde{E} \in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[r(x)] = \max_{x \in K_{\mathscr{P}}} r(x)$$

Proof: Combine strong duality with completeness

Convergence of the SoS hierarchy

Convergence: Let $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfy the Archimedean Assumption

$$\lim_{d \to \infty} \max_{\tilde{E} \in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[r(x)] = \max_{x \in K_{\mathscr{P}}} r(x)$$

Proof: Combine strong duality with completeness

Strong Duality: For all $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfying the Archimedean Assumption

$$\min_{\lambda - r(x) = \Sigma p(x) q_p^2(x)} \lambda = \max_{\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[p]$$

Convergence of the SoS hierarchy

Convergence: Let $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfy the Archimedean Assumption

$$\lim_{d\to\infty} \max_{\tilde{E}\in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[r(x)] = \max_{x\in K_{\mathscr{P}}} r(x)$$

Proof: Combine strong duality with completeness

Strong Duality: For all $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfying the Archimedean Assumption

$$\min_{\lambda - r(x) = \Sigma p(x) q_p^2(x)} \lambda = \max_{\tilde{\mathbb{E}}\in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[p]$$

Putinar's Positivstellensatz: Let $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfy the Archimedean assumption. Then $r(x) > 0$ for all $x \in K_{\mathscr{P}}$ iff

$$r(x) = \sum_{p\in\mathscr{P}\cup\{1\}} p(x) q_p^2(x)$$

Convergence of the SoS hierarchy

Convergence: Let $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfy the Archimedean Assumption

$$\lim_{d\to\infty} \max_{\tilde{E}\in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[r(x)] = \max_{x\in K_{\mathscr{P}}} r(x)$$

When can we guarantee faster convergence?

— Inclusion of axioms such as

- $x_i^2 - x_i = 0 \ \forall i \in [n]$ (hypercube), or
- $1 - x_i^2 = 0 \ \forall i \in [n]$ (hypersphere)

guarantee convergence in degree $2n + deg(\mathscr{P})$

Strong Duality

Strong Duality: For all $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfying the Archimedean Assumption
$$\min_{\lambda - r(x) = \Sigma p(x) q_p^2(x)} \lambda = \max_{\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[p]$$

Idea:
1. Write dual as an SDP searching for the coefficients in the proof.
2. Use SDP strong duality.

Strong Duality

Strong Duality: For all $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfying the Archimedean Assumption $$\min_{\lambda - r(x) = \Sigma p(x) q_p^2(x)} \lambda = \max_{\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[p]$$

Idea:
1. Write dual as an SDP searching for the coefficients in the proof.
2. Use SDP strong duality.

PSD Matrices $Z \in \mathbb{R}^{n^d \times n^d}$ define square polynomials:

By Cholesky Decomposition: $Z = UU^T$

Then $v_d^T U U^T v_d = (v_d^T U)^2 = q^2(x)$. Where $(v_d)_I = \Pi_{i \in I} x_i$

Strong Duality

Strong Duality: For all $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfying the Archimedean Assumption $\min_{\lambda - r(x) = \Sigma p(x) q_p^2(x)} \lambda = \max_{\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[p]$

Idea:

1. Write dual as an SDP searching for the coefficients in the proof.
2. Use SDP strong duality.

PSD Matrices $Z \in \mathbb{R}^{n^d \times n^d}$ define square polynomials:

By Cholesky Decomposition: $Z = UU^T$

Then $v_d^T U U^T v_d = (v_d^T U)^2 = q^2(x)$. Where $(v_d)_I = \Pi_{i \in I} x_i$

Rephrase $\lambda - r(x) = \Sigma_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$ as

$$\begin{aligned} \min \quad & \lambda \\ \text{s.t.} \quad & \lambda - r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) v_{d_p}^T Z_p v_{d_p} & & d_p := (d - deg(p))/2 \\ & Z_p \succeq 0 & & \forall p \in \mathscr{P} \end{aligned}$$

Strong Duality

Strong Duality: For all $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfying the Archimedean Assumption $\min_{\lambda - r(x) = \Sigma p(x) q_p^2(x)} \lambda = \max_{\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[p]$

Rephrase $\lambda - r(x) = \Sigma_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$ as

$$
\begin{aligned}
\min \quad & \lambda \\
\text{s.t.} \quad & \lambda - r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) v_{d_p}^T Z_p v_{d_p} && d_p := (d - deg(p))/2 \\
& Z_p \succeq 0 && \forall p \in \mathscr{P}
\end{aligned}
$$

Strong Duality

Strong Duality: For all $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfying the Archimedean Assumption $\min_{\lambda - r(x) = \Sigma p(x) q_p^2(x)} \lambda = \max_{\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[p]$

Rephrase $\lambda - r(x) = \Sigma_{p \in \mathscr{P} \cup \{1\}} p(x) q_p^2(x)$ as

$$\begin{aligned} \min \quad & \lambda \\ \text{s.t.} \quad & \lambda - r(x) = \sum_{p \in \mathscr{P} \cup \{1\}} p(x) v_{d_p}^T Z_p v_{d_p} && d_p := (d - deg(p))/2 \\ & Z_p \succeq 0 && \forall p \in \mathscr{P} \end{aligned}$$

Removing x variables, this becomes

$$\begin{aligned} \min \quad & \lambda \\ \text{s.t.} \quad & \lambda 1_{[I = \varnothing]} - \vec{r}_I = \sum_{p \in \mathscr{P} \cup \{1\}} \sum_{S+T+K=I} \vec{p}_K (Z_p)_{S,T} && \forall |I| \leq deg(r) \\ & Z_p \succeq 0 && \forall p \in \mathscr{P} \end{aligned}$$

Strong Duality

Strong Duality: For all $\mathscr{P} \subseteq \mathbb{R}[x]$ satisfying the Archimedean Assumption $\min_{\lambda - r(x) = \Sigma p(x) q_p^2(x)} \lambda = \max_{\tilde{\mathbb{E}} \in SOS_d(\mathscr{P})} \tilde{\mathbb{E}}[p]$

Dual:

$$\begin{aligned} \min \quad & \lambda \\ \text{s.t.} \quad & \lambda 1_{[I=\emptyset]} - \vec{r}_I = \sum_{p \in \mathscr{P} \cup \{1\}} \sum_{S+T+K=I} \vec{p}_K (Z_p)_{S,T} \quad \forall |I| \le deg(r) \\ & Z_p \succeq 0 \quad \forall p \in \mathscr{P} \end{aligned}$$

Primal:

$$\begin{aligned} \max \quad & \tilde{\mathbb{E}}[p(x)] \\ \text{s.t.} \quad & M_d \succeq 0 \\ & M_d^p \succeq 0 \quad \forall p \in \mathscr{P} \\ & \tilde{\mathbb{E}}[1] = 1 \end{aligned}$$

Strong duality follows by the SDP strong duality theorem

Outline

1. Developing the Sum-of-Squares Relaxation
2. Phrasing the Relaxation as an SDP
3. The Dual Sum-of-Squares Proofs and Completeness
4. Convergence and Strong Duality
5. Upper Bounds
6. Lower Bounds

Automatizability

Can we find a Sum-of-Squares proof efficiently if it exists?

Claimed: One can find a degree-d Sum-of-Squares proof in time $|\mathscr{P}| \cdot n^{O(d)}$ if it exists.

Automatizability

Can we find a Sum-of-Squares proof efficiently if it exists?

Claimed: One can find a degree-d Sum-of-Squares proof in time $|\mathscr{P}| \cdot n^{O(d)}$ if it exists.

Reasoning: SoS dual is an $|\mathscr{P}| \cdot n^{O(d)}$-size SDP. Can be solved in time $|\mathscr{P}| \cdot n^{O(d)}$ by the Ellipsoid Method (up to additive error ε).

Automatizability

Can we find a Sum-of-Squares proof efficiently if it exists?

Claimed: One can find a degree-d Sum-of-Squares proof in time $|\mathscr{P}| \cdot n^{O(d)}$ if it exists.

Reasoning: SoS dual is an $|\mathscr{P}| \cdot n^{O(d)}$-size SDP. Can be solved in time $|\mathscr{P}| \cdot n^{O(d)}$ by the Ellipsoid Method (up to additive error ε).

This claim is not known to be true in general

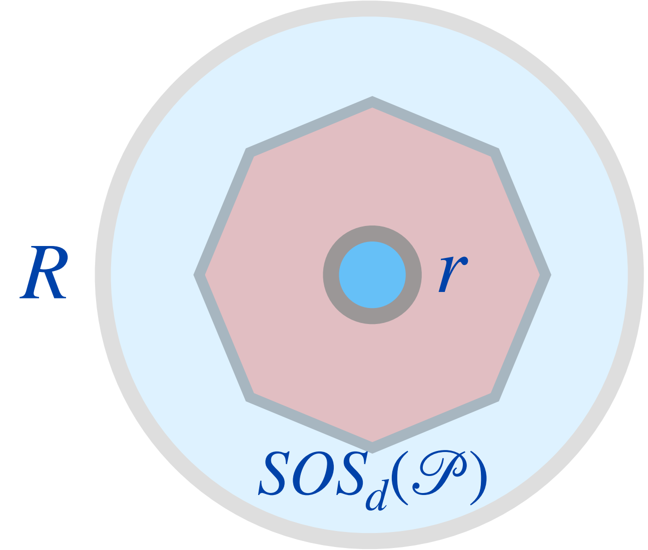
—Even for $\mathscr{P}$ satisfying the Archimedean assumption.

—Even for $\mathscr{P}$ containing $x_i^2 - x_i = 0$ for all $i \in [n]$

Automatizability

Issue:

- Ellipsoid Method requires the feasible set of the SDP to be contained within a ball of radius $R = |\mathscr{P}| \cdot n^{O(d)}$
- i.e. there must exist a proof with bit size $|\mathscr{P}| \cdot n^{O(d)}$



Ellipsoid Method: Let C be a convex set with a polynomial-time separation oracle. For $r, R > 0$ and $c \in \mathbb{R}^n$ such that $Ball(c, r) \subseteq C \subseteq Ball(0,R)$, maximizing over C to an additive error $\varepsilon > 0$ can be done in time $poly(|C|) \cdot \log(R/r\varepsilon)$.

Automatizability

Issue:

- Ellipsoid Method requires the feasible set of the SDP to be contained within a ball of radius $R = |\mathscr{P}| \cdot n^{O(d)}$
- i.e. there must exist a proof with bit size $|\mathscr{P}| \cdot n^{O(d)}$

[RW17] Extending [O'Do17]: There exists small, degree 2 polynomials $\mathscr{P}$, $r(x)$ such that
$-r(x)$ has a degree-2 SoS proof from $\mathscr{P}$,
$-r(x)$ does not admit a degree $o(\sqrt{n})$ proof of polynomial bit length from $\mathscr{P}$.

Automatizability

Issue:

- Ellipsoid Method requires the feasible set of the SDP to be contained within a ball of radius $R = |\mathscr{P}| \cdot n^{O(d)}$
- i.e. there must exist a proof with bit size $|\mathscr{P}| \cdot n^{O(d)}$

[RW17] Extending [O'Do17]: There exists small, degree 2 polynomials $\mathscr{P}$, $r(x)$ such that
$-r(x)$ has a degree-2 SoS proof from $\mathscr{P}$,
$-r(x)$ does not admit a degree $o(\sqrt{n})$ proof of polynomial bit length from $\mathscr{P}$.

Good News: [RW17] provide a set of sufficient conditions under which SoS derivations can be found in time $|\mathscr{P}| \cdot n^{O(d)}$.
—MaxCSP, MaxClique, Balanced Separator, MaxBisection

Automatizability

What about size automatizability?

Automatizability

What about size automatizability?

Open Problem: Is there an algorithm for finding a SoS proof of size s in time $poly(s)$?

Automatizability

What about size automatizability?

Open Problem: Is there an algorithm for finding a SoS proof of size s in time $poly(s)$?

Monomial Size: s_m the minimum number of monomials in any SoS proof.

Size-degree tradeoff [AH18]: Any SoS derivation of monomial size s_m from $\mathscr{P}$ implies a derivation of degree $O(\sqrt{n \log s_m} + deg(\mathscr{P}))$

Automatizability

What about size automatizability?

Open Problem: Is there an algorithm for finding a SoS proof of size s in time $poly(s)$?

Monomial Size: s_m the minimum number of monomials in any SoS proof.

Size-degree tradeoff [AH18]: Any SoS derivation of monomial size s_m from $\mathscr{P}$ implies a derivation of degree $O(\sqrt{n \log s_m} + deg(\mathscr{P}))$

Any SoS derivation of monomial size s_m from a set $\mathscr{P}$ satisfying the conditions of [RW17] can be found in time $n^{O\left(\sqrt{n \log s_m} + deg(\mathscr{P})\right)}$.

Upper Bounds via Sum-of-Squares

Upper bounds leverage strong duality and the $n^{O(d)}$-time SoS algorithm to transform certificates that a solution exists into algorithms for finding that solution.
— Combined with clever rounding schemes

Upper Bounds via Sum-of-Squares

Upper bounds leverage strong duality and the $n^{O(d)}$-time SoS algorithm to transform certificates that a solution exists into algorithms for finding that solution.
— Combined with clever rounding schemes

[GW94]: Degree-2 SoS achieves a 0.878 approximation for MaxCut.
—introduced the random hyperplane rounding technique

Upper Bounds via Sum-of-Squares

Upper bounds leverage strong duality and the $n^{O(d)}$-time SoS algorithm to transform certificates that a solution exists into algorithms for finding that solution.
— Combined with clever rounding schemes

[GW94]: Degree-2 SoS achieves a 0.878 approximation for MaxCut.
—introduced the random hyperplane rounding technique

[ARV04]: Uses degree-4 SoS to obtain a $O(\sqrt{\log n})$-approximation for the Sparsest Cut.
— first use of higher-order SoS relaxations.

Upper Bounds via Sum-of-Squares

Upper bounds leverage strong duality and the $n^{O(d)}$-time SoS algorithm to transform certificates that a solution exists into algorithms for finding that solution.
— Combined with clever rounding schemes

[GW94]: Degree-2 SoS achieves a 0.878 approximation for MaxCut.
—introduced the random hyperplane rounding technique

[ARV04]: Uses degree-4 SoS to obtain a $O(\sqrt{\log n})$-approximation for the Sparsest Cut.
— first use of higher-order SoS relaxations.

A line of work beginning with [KKMO07] uncovered a deep connection between SoS and the Unique Games Conjecture

Upper Bounds via Sum-of-Squares

Upper bounds leverage strong duality and the $n^{O(d)}$-time SoS algorithm to transform certificates that a solution exists into algorithms for finding that solution.
— Combined with clever rounding schemes

[GW94]: Degree-2 SoS achieves a 0.878 approximation for MaxCut.
—introduced the random hyperplane rounding technique

[ARV04]: Uses degree-4 SoS to obtain a $O(\sqrt{\log n})$-approximation for the Sparsest Cut.
— first use of higher-order SoS relaxations.

A line of work beginning with [KKMO07] uncovered a deep connection between SoS and the Unique Games Conjecture

[Rag08]: Assuming the Unique Games Conjecture, degree-2 SoS gives the optimal approximation ratio for every CSP.
— Does not tell us what this approximation ratio is.

Upper Bounds via Sum-of-Squares

[ABS10,BRS11,GS11]: Subexponential-time algorithm for Unique Games based on SoS.

—[BRS11,GS11] Introduced the global correlation rounding technique.

Upper Bounds via Sum-of-Squares

[ABS10,BRS11,GS11]: Subexponential-time algorithm for Unique Games based on SoS.
—[BRS11,GS11] Introduced the global correlation rounding technique.

Global Correlation Rounding:

- Given a pseudo-expectation $\tilde{\mathbb{E}}$, one way to round it is to assign each variable $x_i = 1$ with probability $\tilde{\mathbb{E}}[x_i]$. This can result in poor solutions due to correlations.
- Global Correlation Rounding: for 2CSPs, in expectation, global correlation drops under conditioning on the outcome of a set of random variables, while the objective value remains the same.

Upper Bounds via Sum-of-Squares

[ABS10,BRS11,GS11]: Subexponential-time algorithm for Unique Games based on SoS.
—[BRS11,GS11] Introduced the global correlation rounding technique.

Global Correlation Rounding:

- Given a pseudo-expectation $\tilde{\mathbb{E}}$, one way to round it is to assign each variable $x_i = 1$ with probability $\tilde{\mathbb{E}}[x_i]$. This can result in poor solutions due to correlations.
- Global Correlation Rounding: for 2CSPs, in expectation, global correlation drops under conditioning on the outcome of a set of random variables, while the objective value remains the same.

[BKS13, BKS17]: Developed new rounding techniques for high-dimensional SoS
— Obtained algorithms for problems in quantum information theory, such as Best Separable State.

Average-Case Upper Bounds

Recently, lots of work on average-case algorithms using SoS
—Partly due to an average-case rounding framework introduced in [BKS14]

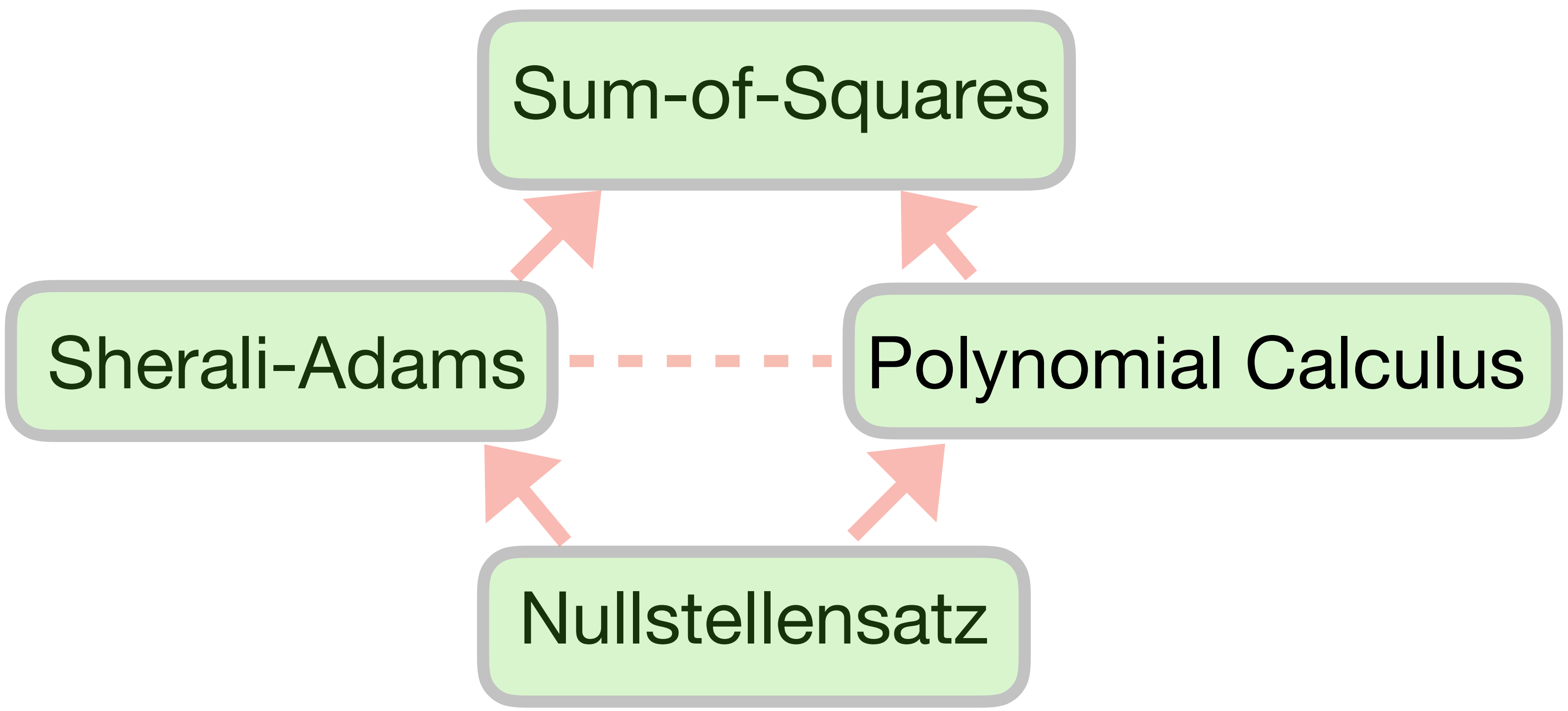
Led to SoS-based algorithms for average-case problems including:
—Dictionary Learning [BKS14],
—Tensor Completion [BM16, PS16],
—Clustering Mixture Models [HL18, KS17],
—Outlier Robust Moment Estimation [KS17],
—Robust Linear Regression [KKM18],
—Attacking cryptographic PRGs [BBKK18, BHKS19].

Outline

1. Developing the Sum-of-Squares Relaxation
2. Phrasing the Relaxation as an SDP
3. The Dual Sum-of-Squares Proofs and Completeness
4. Convergence and Strong Duality
5. Upper Bounds
6. Lower Bounds

Comparison with other Proof Systems

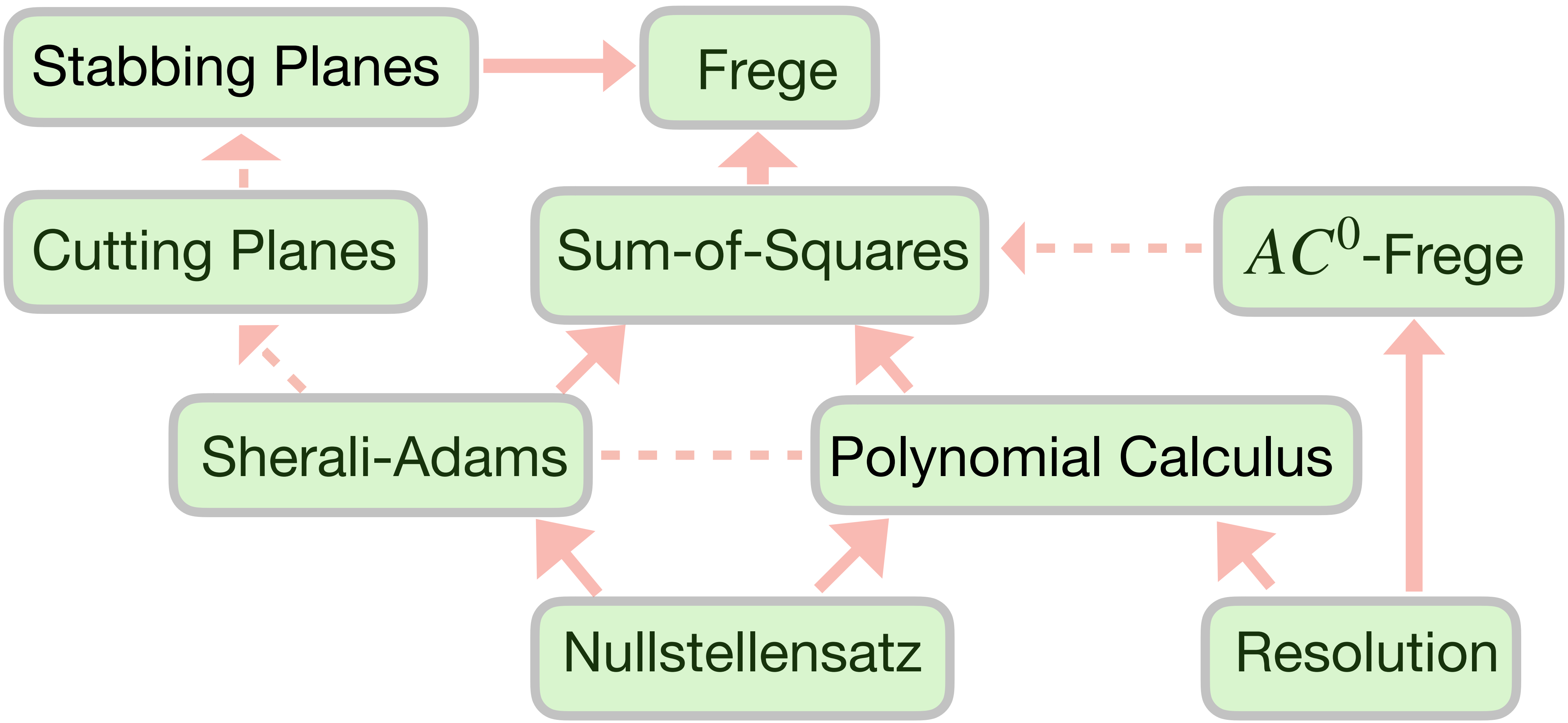
Simulations in terms of degree



Many of these separations as well as the simulation of PC by SoS are due to [Ber18]

Comparison with other Proof Systems

Simulation in terms of size



Open Questions:

— Does SoS simulate AC^0-Frege?

— How does SoS compare to Cutting Planes?

— How does SoS compare to Stabbing Planes / R(CP)?

Lower Bounds on SoS

If degree-d SoS cannot refute $\mathscr{P} \cup \{r(x) - \lambda\}$ then maximizing $r(x)$ over the degree-d SoS relaxation of $\mathscr{P}$ attains a value of at least λ.

— Lower bounds on the degree of SoS refutations imply inapproximability results for the SoS hierarchy.

Lower Bounds on SoS

If degree-d SoS cannot refute $\mathscr{P} \cup \{r(x) - \lambda\}$ then maximizing $r(x)$ over the degree-d SoS relaxation of $\mathscr{P}$ attains a value of at least λ.

— Lower bounds on the degree of SoS refutations imply inapproximability results for the SoS hierarchy.

To prove a degree lower bound of d on refuting a set of polynomials $\mathscr{P}$, one constructs a degree-d pseudo-expectation for $\mathscr{P}$

Lower Bounds on SoS

If degree-d SoS cannot refute $\mathscr{P} \cup \{r(x) - \lambda\}$ then maximizing $r(x)$ over the degree-d SoS relaxation of $\mathscr{P}$ attains a value of at least λ.

— Lower bounds on the degree of SoS refutations imply inapproximability results for the SoS hierarchy.

To prove a degree lower bound of d on refuting a set of polynomials $\mathscr{P}$, one constructs a degree-d pseudo-expectation for $\mathscr{P}$

Random 3XOR: [Gri01, Sch08] systems of random 3XOR equations require degree $\Omega(n)$.

- Reduction to Resolution width lower bounds.
- Builds on earlier ideas [BGIP01, Gri98] for NS and PC.
- [Sch08] Implies lower bounds on Max3SAT, Max Ind Set.

Lower Bounds on SoS — CSPs

Approximation Resistant: The best polynomial-time approximation is a uniformly random assignment.

Lower Bounds on SoS — CSPs

Approximation Resistant: The best polynomial-time approximation is a uniformly random assignment.

[Chan13]: Assuming $P \neq NP$, any predicate $P : \{0,1\}^k \to \{0,1\}$ that is pairwise independent and algebraically linear is approximation resistant

- Pairwise Independent: $P^{-1}(1)$ supports a distribution μ such that the pairwise marginals $\mu_i \mu_j$ for $i \neq j$ is uniform over $\{0,1\}^2$.
- Algebraically Linear: μ is also the uniform distribution over a subspace $V \subseteq GF(2)$.

Lower Bounds on SoS — CSPs

Approximation Resistant: The best polynomial-time approximation is a uniformly random assignment.

[Chan13]: Assuming $P \neq NP$, any predicate $P : \{0,1\}^k \to \{0,1\}$ that is pairwise independent and algebraically linear is approximation resistant

- Pairwise Independent: $P^{-1}(1)$ supports a distribution μ such that the pairwise marginals $\mu_i \mu_j$ for $i \neq j$ is uniform over $\{0,1\}^2$.
- Algebraically Linear: μ is also the uniform distribution over a subspace $V \subseteq GF(2)$.

[AM09]: Assuming the UGC, any predicate $P : \{0,1\}^k \to \{0,1\}$ that is pairwise uniform is approximation resistant

Lower Bounds on SoS — CSPs

Approximation Resistant CSP for Degree-d SoS: If there is an instance such that

- A random assignment is essentially optimal
- Degree-d SoS believes $1 - o(1)$ fraction of constraints can be satisfied

Lower Bounds on SoS — CSPs

Approximation Resistant CSP for Degree-d SoS: If there is an instance such that

- A random assignment is essentially optimal
- Degree-d SoS believes $1 - o(1)$ fraction of constraints can be satisfied

[Tul09]: Any CSP on pairwise uniform and algebraically linear predicates is approximation resistant for degree $\Omega(n)$ SoS

- Method for doing reductions in SoS
- Lower bounds for problems such as Vertex Cover, IndSet

Open Question: Prove that SoS cannot achieve better than a 2-approximation for Vertex Cover.

Lower Bounds on SoS — CSPs

Approximation Resistant CSP for Degree-d SoS: If there is an instance such that

- A random assignment is essentially optimal
- Degree-d SoS believes $1 - o(1)$ fraction of constraints can be satisfied

[Tul09]: Any CSP on pairwise uniform and algebraically linear predicates is approximation resistant for degree $\Omega(n)$ SoS

- Method for doing reductions in SoS
- Lower bounds for problems such as Vertex Cover, IndSet

[BCK15]: Any CSP defined on pairwise uniform predicates is approximation resistant for degree $\Omega(n)$ SoS

Open Question: Prove that SoS cannot achieve better than a 2-approximation for Vertex Cover.

Average-Case Lower Bounds

Random CSPs: [KMOW17] Proved sharp lower bounds that tightly characterize the number of clauses needed for SoS to refute random CSP instances with a given predicate P.
— Matches the upper bounds of [AOW15, RRS16].

Average-Case Lower Bounds

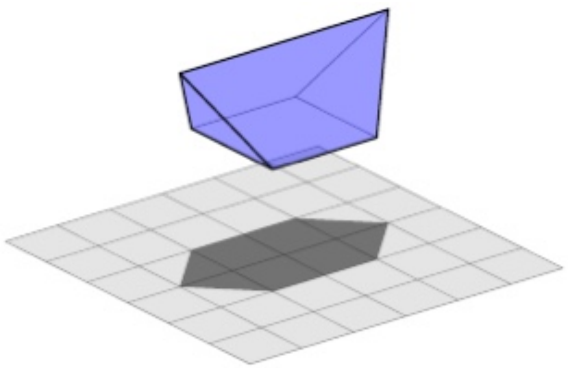
Random CSPs: [KMOW17] Proved sharp lower bounds that tightly characterize the number of clauses needed for SoS to refute random CSP instances with a given predicate P.
— Matches the upper bounds of [AOW15, RRS16].

Planted Clique: [MPW15, HKPRS18], culminating in [BHKKMP18] proved nearly tight lower bounds on the degree of SoS proofs of the Planted Clique problem.
— Introduced the pseudo-calibration framework; a computational bayesian approach to constructing pseudo-expectations.

Applications of Lower Bounds

(SDP) Extended Formulation: Of a polytope P is any polytope (spectahedron) Q such that there exists a linear projection such that $proj(Q) = P$.

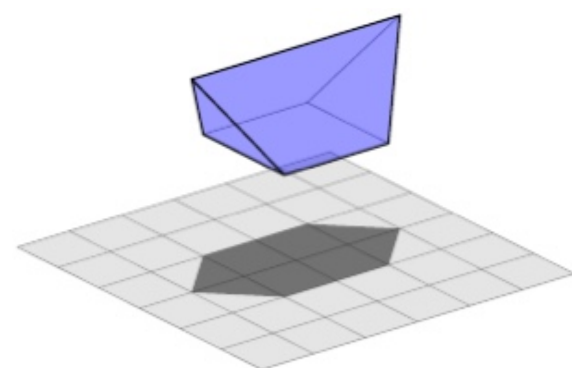
— Restriction: Polytope is instance independent



Applications of Lower Bounds

(SDP) Extended Formulation: Of a polytope P is any polytope (spectahedron) Q such that there exists a linear projection such that $proj(Q) = P$.

—Restriction: Polytope is instance independent

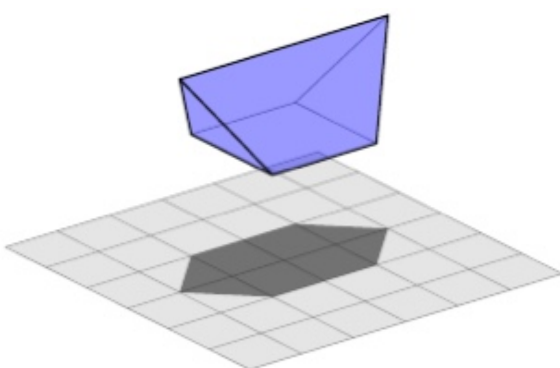


[CLRS13, KMR17] For any CSP on $N = n$ variables, there is a constant c such that no size $n^{c \cdot d}$ extended formulation can achieve a better approximation than degree-d Sherali-Adams.

Applications of Lower Bounds

(SDP) Extended Formulation: Of a polytope P is any polytope (spectahedron) Q such that there exists a linear projection such that $proj(Q) = P$.

—Restriction: Polytope is instance independent



[CLRS13, KMR17] For any CSP on $N = n$ variables, there is a constant c such that no size $n^{c \cdot d}$ extended formulation can achieve a better approximation than degree-d Sherali-Adams.

[LRS14] For any CSP, there exists a constant c such that no size $c(n/\log n)^{d/4}$ SDP extended formulation can achieve a better approximation on any instance of $N = n^{4d}$ variables than degree-d Sum-of-Squares can on n variables.

Thank You!