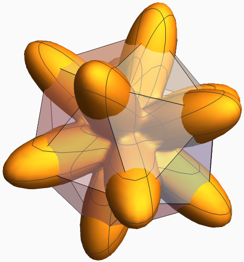
# Semidefinite Relaxations of Products of Nonnegative Forms

Workshop on Real Algebraic Geometry and Algorithms for Geometric Constraint Systems

Chenyang Yuan (joint work with Pablo Parrilo)

June 18, 2021



How to exploit product structure in polynomial optimization problems

# Introduction

How to exploit product structure in polynomial optimization problems

⇓

Computational tractability

How to exploit product structure in polynomial optimization problems

$$\Downarrow$$

Computational tractability

$$+$$

Provable approximation guarantees

# Introduction

Given $\mathcal{A} = (A_1, \ldots, A_d)$ where $A_i \succeq 0$, we study the following polynomial optimization problem on $\mathbb{K}^n = \mathbb{R}^n$ or $\mathbb{C}^n$:

$$\textsc{Opt}(\mathcal{A}) := \max_{x \in \mathbb{K}^n,\, \|x\|=1} \left( \prod_{i=1}^{d} \langle x, A_i x \rangle \right)^{1/d}$$

# Introduction

Given $\mathcal{A} = (A_1, \ldots, A_d)$ where $A_i \succeq 0$, we study the following polynomial optimization problem on $\mathbb{K}^n = \mathbb{R}^n$ or $\mathbb{C}^n$:

$$\mathrm{OPT}(\mathcal{A}) := \max_{x \in \mathbb{K}^n,\, \|x\|=1} \left( \prod_{i=1}^{d} \langle x, A_i x \rangle \right)^{1/d}$$

Properties:

- High-degree polynomial optimization problem: degree $2d$ in $n$ variables

# Introduction

Given $\mathcal{A} = (A_1, \ldots, A_d)$ where $A_i \succeq 0$, we study the following polynomial optimization problem on $\mathbb{K}^n = \mathbb{R}^n$ or $\mathbb{C}^n$:

$$\mathrm{OPT}(\mathcal{A}) := \max_{x \in \mathbb{K}^n,\, \|x\|=1} \left( \prod_{i=1}^{d} \langle x, A_i x \rangle \right)^{1/d}$$

Properties:

- High-degree polynomial optimization problem: degree $2d$ in $n$ variables
- Compact representation: represented in $O(n^2 d)$ space

# Introduction

Given $\mathcal{A} = (A_1, \ldots, A_d)$ where $A_i \succeq 0$, we study the following polynomial optimization problem on $\mathbb{K}^n = \mathbb{R}^n$ or $\mathbb{C}^n$:

$$\textsc{Opt}(\mathcal{A}) := \max_{x \in \mathbb{K}^n,\, \|x\|=1} \left( \prod_{i=1}^{d} \langle x, A_i x \rangle \right)^{1/d}$$

Properties:

- High-degree polynomial optimization problem: degree $2d$ in $n$ variables
- Compact representation: represented in $O(n^2 d)$ space
- Polynomial time solution when $d$ is fixed, NP-hard when $d = \Omega(n)$

# Applications and Motivation

$$\textsc{Opt}(\mathcal{A}) := \max_{\|x\|=1} \prod_{i=1}^{d} \langle x, A_i x \rangle^{1/d}$$

# Applications and Motivation

$$\textsc{Opt}(\mathcal{A}) \coloneqq \max_{\|x\|=1} \prod_{i=1}^{d} \langle x, A_i x \rangle^{1/d}$$

Kantorovich's inequality ($A_1 = Q$, $A_2 = Q^{-1}$): Given $Q \succ 0$,

$$\max_{\|x\|=1} \sqrt{(x^T Q x)(x^T Q^{-1} x)} \leq \frac{1}{4} \left( \sqrt{\frac{\lambda_1(Q)}{\lambda_n(Q)}} + \sqrt{\frac{\lambda_n(Q)}{\lambda_1(Q)}} \right)^2$$

# Applications and Motivation

$$\textsc{Opt}(\mathcal{A}) \coloneqq \max_{\|x\|=1} \prod_{i=1}^{d} \langle x, A_i x\rangle^{1/d}$$

Kantorovich's inequality ($A_1 = Q$, $A_2 = Q^{-1}$): Given $Q \succ 0$,

$$\max_{\|x\|=1} \sqrt{(x^T Q x)(x^T Q^{-1} x)} \le \frac{1}{4}\left(\sqrt{\frac{\lambda_1(Q)}{\lambda_n(Q)}} + \sqrt{\frac{\lambda_n(Q)}{\lambda_1(Q)}}\right)^2$$

Approximating permanents of PSD matrices [YP20] ($A_i = v_i v_i^\dagger$): Let $M = V^\dagger V$, $v_i$ columns of $V$.

$$r(M) \coloneqq \max_{\|x\|=1, x\in\mathbb{C}^n} \prod_{i=1}^{n} |\langle x, v_i\rangle|^2, \quad \frac{n!}{n^n} r(M) \le \mathrm{per}(M)$$

# Applications and Motivation

$$\textsc{Opt}(\mathcal{A}) \coloneqq \max_{\|x\|=1} \prod_{i=1}^{d} \langle x, A_i x\rangle^{1/d}$$

Kantorovich's inequality ($A_1 = Q$, $A_2 = Q^{-1}$): Given $Q \succ 0$,

$$\max_{\|x\|=1} \sqrt{(x^T Q x)(x^T Q^{-1} x)} \leq \frac{1}{4}\left(\sqrt{\frac{\lambda_1(Q)}{\lambda_n(Q)}} + \sqrt{\frac{\lambda_n(Q)}{\lambda_1(Q)}}\right)^2$$

Approximating permanents of PSD matrices [YP20] ($A_i = v_i v_i^\dagger$): Let $M = V^\dagger V$, $v_i$ columns of $V$.

$$r(M) \coloneqq \max_{\|x\|=1, x\in\mathbb{C}^n} \prod_{i=1}^{n} \left|\langle x, v_i\rangle\right|^2, \quad \frac{n!}{n^n} r(M) \leq \mathrm{per}(M)$$

Portfolio optimization ($A_i = \mathrm{diag}(r_i)$): Given rates of return over a time period $r_1, \ldots, r_T \in \mathbb{R}_+^n$, maximize expected profit:

$$\max_{y\geq 0, \sum_i y_i = 1} \left(\prod_{i=1}^{T} \langle y, r_i\rangle\right)^{1/d}$$

# Applications and Motivation

$$\textsc{Opt}(\mathcal{A}) \coloneqq \max_{\|x\|=1} \prod_{i=1}^{d} \langle x, A_i x\rangle^{1/d}$$

Kantorovich's inequality ($A_1 = Q$, $A_2 = Q^{-1}$): Given $Q \succ 0$,

$$\max_{\|x\|=1} \sqrt{(x^T Q x)(x^T Q^{-1} x)} \leq \frac{1}{4}\left(\sqrt{\frac{\lambda_1(Q)}{\lambda_n(Q)}} + \sqrt{\frac{\lambda_n(Q)}{\lambda_1(Q)}}\right)^2$$

Approximating permanents of PSD matrices [YP20] ($A_i = v_i v_i^\dagger$): Let $M = V^\dagger V$, $v_i$ columns of $V$.

$$r(M) \coloneqq \max_{\|x\|=1, x\in\mathbb{C}^n} \prod_{i=1}^{n} |\langle x, v_i\rangle|^2, \quad \frac{n!}{n^n} r(M) \leq \operatorname{per}(M)$$

Portfolio optimization ($A_i = \operatorname{diag}(r_i)$): Given rates of return over a time period $r_1, \dots, r_T \in \mathbb{R}_+^n$, maximize expected profit:

$$\max_{y\geq 0, \sum_i y_i = 1} \left(\prod_{i=1}^{T} \langle y, r_i\rangle\right)^{1/d}$$

And more! (Solving systems of quadratic equations, linear polarization constants, Nash social welfare ...)

# Our Contributions

Using semidefinite programming (SDP) based approximation algorithms for general polynomial optimization (Sum-of-Squares):

Compute: SDP with $O(\binom{n+d}{d})$ vars/consts   Approx: $\Omega(\frac{1}{n})$

# Our Contributions

Using semidefinite programming (SDP) based approximation algorithms for general polynomial optimization (Sum-of-Squares):

Compute: SDP with $O(\binom{n+d}{d})$ vars/consts Approx: $\Omega(\frac{1}{n})$

Our relaxation and approximation algorithm that exploits compact representation:

Compute: SDP with $O(n^2 d)$ vars/consts Approx: $\Omega(1)$

# Our Contributions

Using semidefinite programming (SDP) based approximation algorithms for general polynomial optimization (Sum-of-Squares):

Compute: SDP with $O(\binom{n+d}{d})$ vars/consts Approx: $\Omega(\frac{1}{n})$

Our relaxation and approximation algorithm that exploits compact representation:

Compute: SDP with $O(n^2 d)$ vars/consts Approx: $\Omega(1)$

We also:

- Prove that when $d = \Omega(n)$, NP-hard to approximate

# Our Contributions

Using semidefinite programming (SDP) based approximation algorithms for general polynomial optimization (Sum-of-Squares):

Compute: SDP with $O(\binom{n+d}{d})$ vars/consts Approx: $\Omega(\frac{1}{n})$

Our relaxation and approximation algorithm that exploits compact representation:

Compute: SDP with $O(n^2 d)$ vars/consts Approx: $\Omega(1)$

We also:

- Prove that when $d = \Omega(n)$, NP-hard to approximate
- Introduce higher-degree relaxations that trade off computation with approximation quality

# Our Contributions

Using semidefinite programming (SDP) based approximation algorithms for general polynomial optimization (Sum-of-Squares):

Compute: SDP with $O(\binom{n+d}{d})$ vars/consts  Approx: $\Omega(\frac{1}{n})$

Our relaxation and approximation algorithm that exploits compact representation:

Compute: SDP with $O(n^2 d)$ vars/consts  Approx: $\Omega(1)$

We also:

- Prove that when $d = \Omega(n)$, NP-hard to approximate
- Introduce higher-degree relaxations that trade off computation with approximation quality
- Exhibit integrality gap instances that show our analysis of our SDP based relaxation is tight

# Semidefinite Relaxation

$$\max_{\|x\|=1} \prod_{i=1}^{d} \langle x, A_i x \rangle^{1/d} \qquad \text{(Opt)}$$

rank-1 relax ↓

$$\max_{X} \prod_{i=1}^{d} \langle X, A_i \rangle^{1/d} \quad \text{s.t.} \quad \mathrm{Tr}(X) = 1,\ X \succeq 0$$

use AM-GM ↓

$$\min_{\lambda,\alpha} \lambda \quad \text{s.t.} \quad \frac{1}{d}\sum_{i=1}^{d} \alpha_i A_i \preceq \lambda I, \quad \prod_i \alpha_i \geq 1,\ \alpha_i > 0 \qquad \text{(OptSDP)}$$

dual: the two problems above are connected by ⟷

# Semidefinite Relaxation

$$\max_{\|x\|=1} \prod_{i=1}^{d} \langle x, A_i x \rangle^{1/d} \qquad (\textsc{Opt})$$

rank-1 relax

use AM-GM

$$\max_{X} \prod_{i=1}^{d} \langle X, A_i \rangle^{1/d} \quad \text{s.t. } \operatorname{Tr}(X) = 1,\ X \succeq 0$$

dual $\longleftrightarrow$

$$\min_{\lambda,\alpha} \lambda \quad \text{s.t. } \frac{1}{d}\sum_{i=1}^{d} \alpha_i A_i \preceq \lambda I, \quad \prod_i \alpha_i \geq 1,\ \alpha_i > 0 \qquad (\textsc{OptSDP})$$

**Theorem**
*Let* $r = \operatorname{rank}(X^*) \leq n$, *$\gamma$ be Euler's constant, $\phi$ be digamma function*

$$c_r(\mathbb{K})\, \textsc{OptSDP} \leq \textsc{Opt} \leq \textsc{OptSDP}$$

# Semidefinite Relaxation

$$\max_{\|x\|=1} \prod_{i=1}^{d} \langle x, A_i x\rangle^{1/d} \qquad (\textsc{Opt})$$

rank-1 relax ↓ use AM-GM ↓

$$\begin{aligned} &\max_{X} \prod_{i=1}^{d} \langle X, A_i\rangle^{1/d} \\ &\text{s.t.} \quad \mathrm{Tr}(X) = 1,\ X \succeq 0 \end{aligned} \quad \overset{\text{dual}}{\longleftrightarrow} \quad \begin{aligned} &\min_{\lambda,\alpha} \lambda \\ &\text{s.t.} \quad \frac{1}{d}\sum_{i=1}^{d} \alpha_i A_i \preceq \lambda I \\ &\qquad \prod_i \alpha_i \geq 1,\ \alpha_i > 0 \end{aligned} \qquad (\textsc{OptSDP})$$

**Theorem**
*Let $r = \mathrm{rank}(X^*) \leq n$, $\gamma$ be Euler's constant, $\phi$ be digamma function*

$$c_r(\mathbb{K})\, \textsc{OptSDP} \leq \textsc{Opt} \leq \textsc{OptSDP}$$

$$c_r(\mathbb{K}) = \begin{cases} \exp(-\gamma - \log 2 - \phi\left(\frac{r}{2}\right) + \log\left(\frac{r}{2}\right)) > 0.2807 & \textit{if } \mathbb{K} = \mathbb{R} \\ \exp(-\gamma - \phi(r) + \log(r)) > 0.5614 & \textit{if } \mathbb{K} = \mathbb{C} \end{cases}$$

# Proof Sketch

Rounding algorithm: Given optimum $X^*$, produce unit vector $\hat{x}$ by:

# Proof Sketch

Rounding algorithm: Given optimum $X^*$, produce unit vector $\hat{x}$ by:

- Sample $x \sim N(0, X^*)$
- Normalize $\hat{x} = x / \|x\|$

## Proof Sketch

Rounding algorithm: Given optimum $X^*$, produce unit vector $\hat{x}$ by:

- Sample $x \sim N(0, X^*)$
- Normalize $\hat{x} = x / \|x\|$

Lower bound expected value of objective:

$$\begin{aligned}
\textsc{Opt} &\geq \mathbb{E}\left[\prod_{i=1}^{d} \langle \hat{x}, A_i \hat{x} \rangle^{1/d}\right] \\
&= \mathbb{E}\left[\exp\left(\frac{1}{d}\sum_{i=1}^{d} \log \langle \hat{x}, A_i \hat{x} \rangle\right)\right] \\
&\geq \exp\left(\frac{1}{d}\sum_{i=1}^{d} \mathbb{E}[\log \langle \hat{x}, A_i \hat{x} \rangle]\right) \\
&\geq c_r(\mathbb{K})\textsc{OptSDP}
\end{aligned}$$

# Application: convex hull of image of quadratic map

Let $\varphi(x) : \mathbb{K}^n \to \mathbb{K}^d$ be a quadratic map: $x \mapsto (\langle x, A_1 x \rangle, \ldots, \langle x, A_d x \rangle)$

# Application: convex hull of image of quadratic map

Let $\varphi(x) : \mathbb{K}^n \to \mathbb{K}^d$ be a quadratic map: $x \mapsto (\langle x, A_1 x \rangle, \ldots, \langle x, A_d x \rangle)$

When is $\varphi(\mathbb{K}^n)$ convex? (always true for $d = 2$, not in general)

# Application: convex hull of image of quadratic map

Let $\varphi(x) : \mathbb{K}^n \to \mathbb{K}^d$ be a quadratic map: $x \mapsto (\langle x, A_1 x\rangle, \dots, \langle x, A_d x\rangle)$

When is $\varphi(\mathbb{K}^n)$ convex? (always true for $d = 2$, not in general)

How far is $\varphi(\mathbb{K}^n)$ from $\mathrm{conv}(\varphi(\mathbb{K}^n))$? Measure relative entropy distance between intersection of these cones and the simplex $\Delta_d$

# Application: convex hull of image of quadratic map

Let $\varphi(x) : \mathbb{K}^n \to \mathbb{K}^d$ be a quadratic map: $x \mapsto (\langle x, A_1 x\rangle, \ldots, \langle x, A_d x\rangle)$

When is $\varphi(\mathbb{K}^n)$ convex? (always true for $d = 2$, not in general)

How far is $\varphi(\mathbb{K}^n)$ from $\mathrm{conv}(\varphi(\mathbb{K}^n))$? Measure relative entropy distance between intersection of these cones and the simplex $\Delta_d$

**Theorem**
*Let $a \in \mathrm{conv}(\varphi(\mathbb{K}^n)) \cap \Delta_d$. Then there exists a point $b \in \varphi(\mathbb{K}^n) \cap \Delta_d$ such that*

$$D(a \parallel b) = \sum_{i=1}^{d} a_i \ln\left(\frac{a_i}{b_i}\right) \le \log(c_r(\mathbb{K}))$$

Proved by Barvinok (2014) for a larger constant, our analysis gives asymptotically optimal constant

## Higher-order Relaxations

$\textsc{OptSDP}$ constructed using AM/GM inequality. If $\prod_i \alpha_i = 1$,

$$\prod_{i=1}^{d} \langle x, A_i x\rangle^{1/d} \leq x^T \left( \frac{1}{d} \sum_{i=1}^{d} \alpha_i A_i \right) x$$

Can we get a better bound with a higher-degree polynomial on the RHS?

# Higher-order Relaxations

$\textsc{OptSDP}$ constructed using AM/GM inequality. If $\prod_i \alpha_i = 1$,

$$\prod_{i=1}^{d} \langle x, A_i x \rangle^{1/d} \leq x^T \left( \frac{1}{d} \sum_{i=1}^{d} \alpha_i A_i \right) x$$

Can we get a better bound with a higher-degree polynomial on the RHS?

Let $E_k$ be elementary symmetric polynomials:

$$E_k(y_1, \ldots, y_d) = \binom{d}{k}^{-1} \sum_{I \subseteq [d], |I| = k} \prod_{i \in I} x_i$$

# Higher-order Relaxations

$\textsc{OptSDP}$ constructed using AM/GM inequality. If $\prod_i \alpha_i = 1$,

$$\prod_{i=1}^{d} \langle x, A_i x \rangle^{1/d} \leq x^T \left( \frac{1}{d} \sum_{i=1}^{d} \alpha_i A_i \right) x$$

Can we get a better bound with a higher-degree polynomial on the RHS?

Let $E_k$ be elementary symmetric polynomials:

$$E_k(y_1, \ldots, y_d) = \binom{d}{k}^{-1} \sum_{I \subseteq [d], |I| = k} \prod_{i \in I} x_i$$

Maclaurin's inequality:

$$(y_1 \cdots y_d)^{1/d} = E_d^{1/d} \leq E_{d-1}^{1/(d-1)} \leq \cdots \leq E_2^{1/2} \leq E_1 = \frac{y_1 + \cdots + y_d}{d}$$

# Higher-order Relaxations

$\textsc{OptSDP}$ constructed using AM/GM inequality. If $\prod_i \alpha_i = 1$,

$$\prod_{i=1}^{d} \langle x, A_i x \rangle^{1/d} \leq x^T \left( \frac{1}{d} \sum_{i=1}^{d} \alpha_i A_i \right) x$$

Can we get a better bound with a higher-degree polynomial on the RHS?

Let $E_k$ be elementary symmetric polynomials:

$$E_k(y_1, \ldots, y_d) = \binom{d}{k}^{-1} \sum_{I \subseteq [d], |I| = k} \prod_{i \in I} x_i$$

Maclaurin's inequality:

$$(y_1 \cdots y_d)^{1/d} = E_d^{1/d} \leq E_{d-1}^{1/(d-1)} \leq \cdots \leq E_2^{1/2} \leq E_1 = \frac{y_1 + \cdots + y_d}{d}$$

Use Sum-of-Squares to construct relaxations $\textsc{OptSoS}_k$ for $1 \leq k \leq d$

$$\textsc{Opt} \leq \textsc{OptSDP} = \textsc{OptSoS}_1 \leq \textsc{OptSoS}_d$$

Trades off computation for accuracy

# Example: Icosahedral form

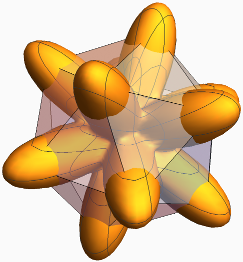
Let $\psi$ be golden ratio, $C$ chosen so that $\max_{x^2+y^2+z^2=1} p(x,y,z) = 1$.

$$p(x,y,z) = C\left[(x+\psi y)(x-\psi y)(y+\psi z)(y-\psi z)(z+\psi x)(z-\psi x)\right]^2$$

# Example: Icosahedral form

Let $\psi$ be golden ratio, $C$ chosen so that $\max_{x^2+y^2+z^2=1} p(x,y,z) = 1$.

$$p(x,y,z) = C\left[(x+\psi y)(x-\psi y)(y+\psi z)(y-\psi z)(z+\psi x)(z-\psi x)\right]^2$$

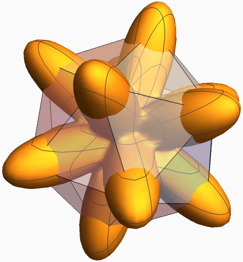


Hard instance because of high degree of symmetry!

# Example: Icosahedral form

Let $\psi$ be golden ratio, $C$ chosen so that $\max_{x^2+y^2+z^2=1} p(x,y,z) = 1$.

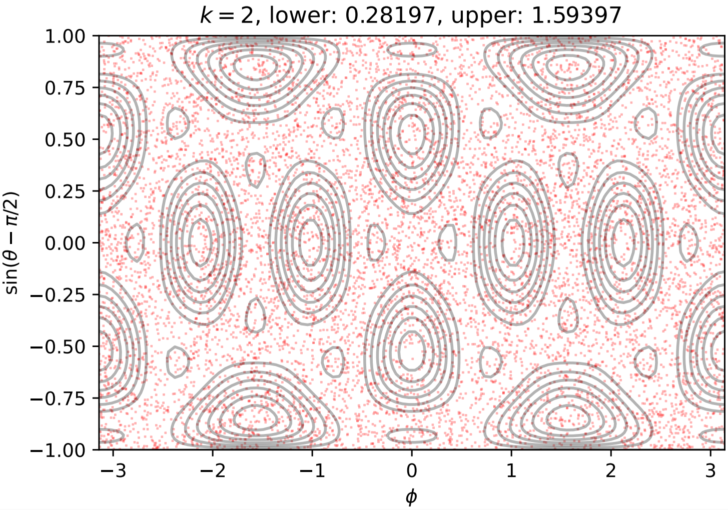
$$p(x,y,z) = C\left[(x+\psi y)(x-\psi y)(y+\psi z)(y-\psi z)(z+\psi x)(z-\psi x)\right]^2$$



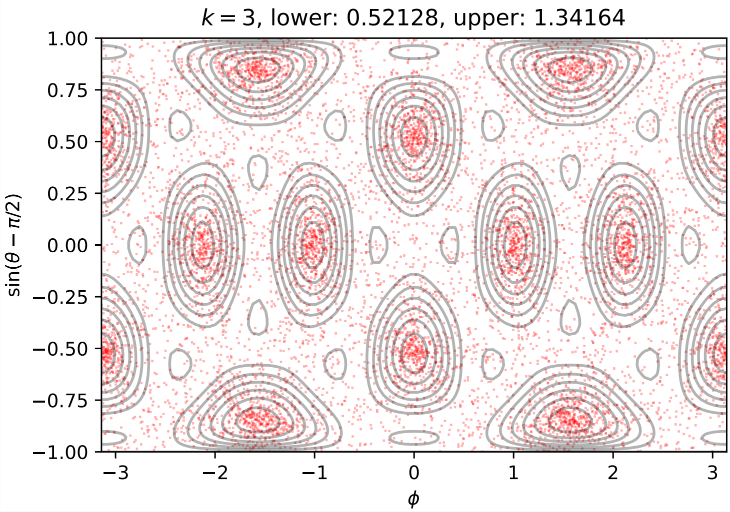
Hard instance because of high degree of symmetry!

We implement a randomized rounding algorithm to obtain feasible solution from relaxations $\textsc{OptSoS}_k$

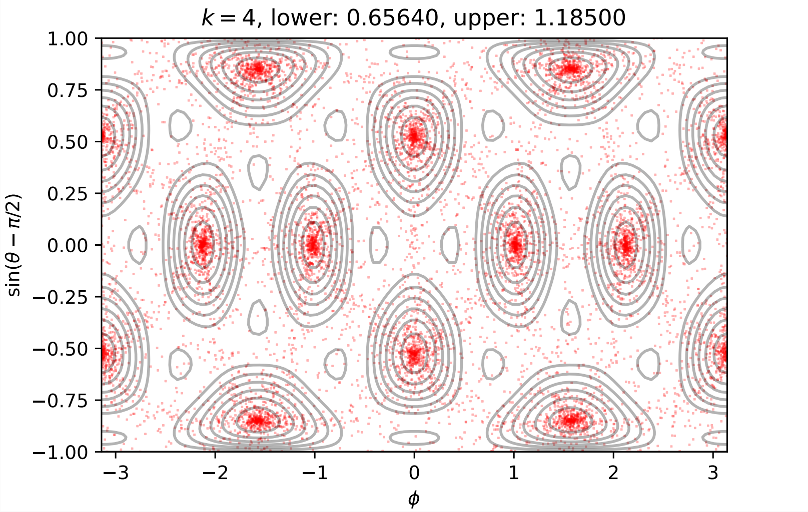
# Distribution sampled from rounding algorithm



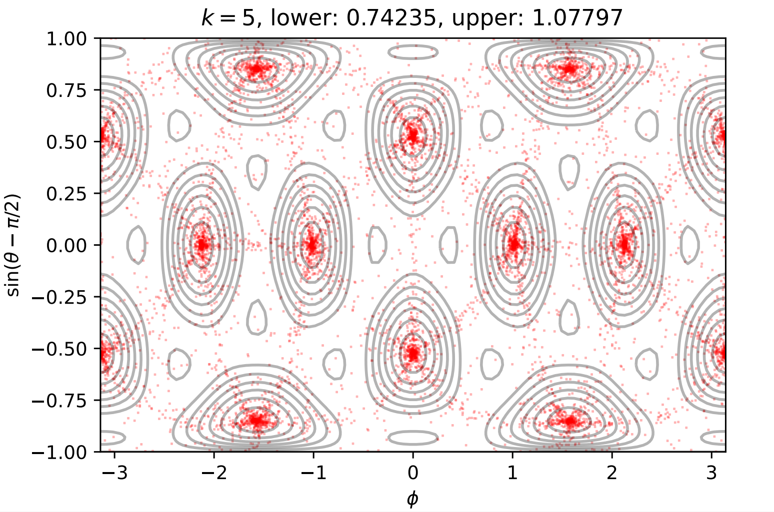
# Distribution sampled from rounding algorithm



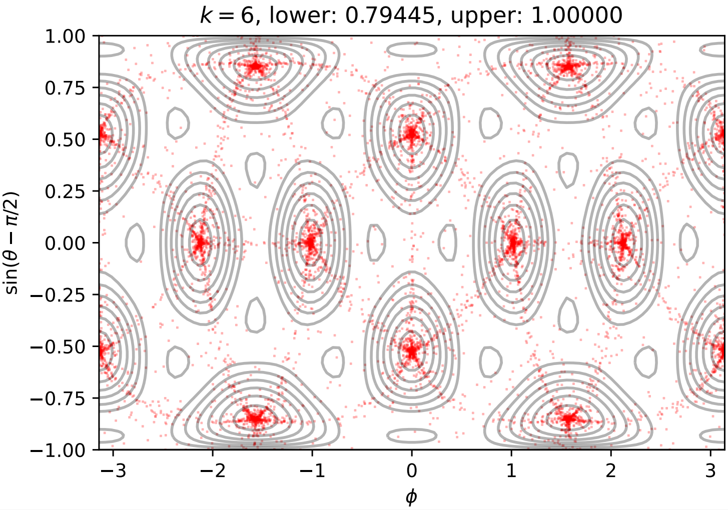
# Distribution sampled from rounding algorithm



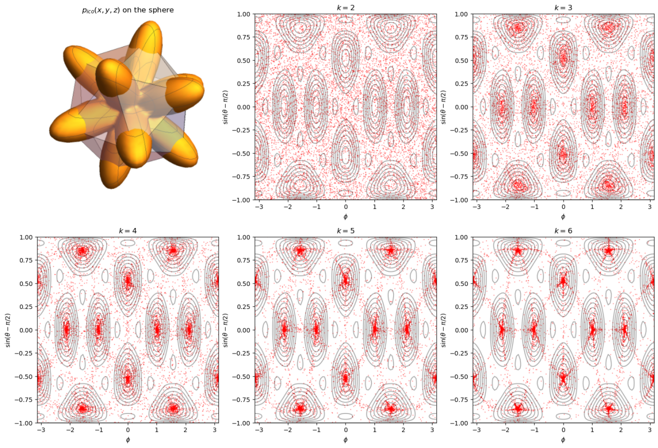
# Distribution sampled from rounding algorithm



# Distribution sampled from rounding algorithm



# Distribution sampled from rounding algorithm




Distribution concentrates towards optima as $k$ increases

# Conclusion

Takeaways:

- Product of PSD forms has nice structure and generalizes many problems
- Exploiting product structure allows us to write a computationally efficient relaxation with good approximation guarantees

# Conclusion

Takeaways:

- Product of PSD forms has nice structure and generalizes many problems
- Exploiting product structure allows us to write a computationally efficient relaxation with good approximation guarantees

Future work:

- Low-rank guarantees of solution from symmetry
- How to generate intermediate Sum-of-Squares relaxations for other high degree polynomial optimization problems?