DSOS/SDOS Programming: New Tools for Optimization over Nonnegative Polynomials

Amir Ali Ahmadi

Princeton University

Dept. of Operations Research and Financial Engineering (ORFE)

Joint work with:

Anirudha Majumdar

MIT, EECS

The Joint Mathematics Meeting of AMS and MAA
January, 2015

PRINCETON UNIVERSITY ORFE

Optimization over nonnegative polynomials

Defn. A polynomial $p(x) := p(x_1, \ldots, x_n)$ is nonnegative if $p(x) \geq 0, \forall x \in \mathbb{R}^n$.

Ex. Decide if the following polynomial is nonnegative:

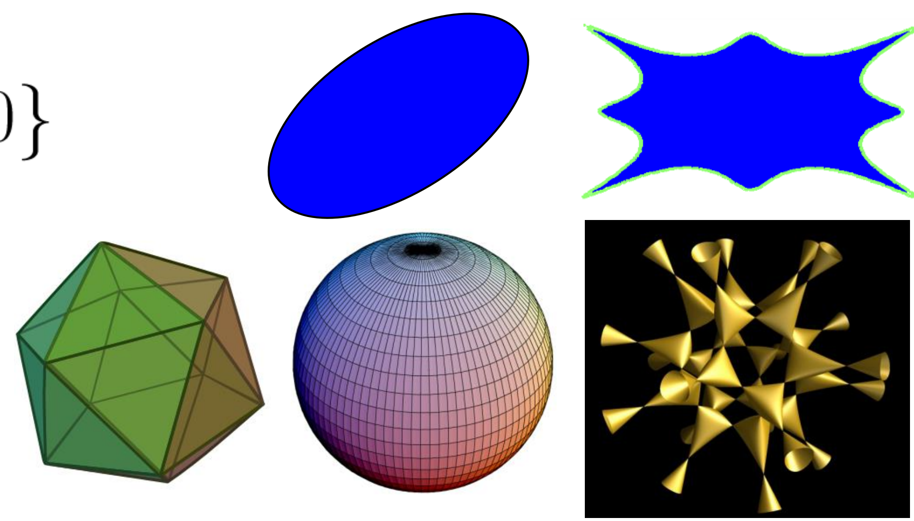
$$
\begin{aligned}
p(x) \quad = \quad & x_1^4 - 6x_1^3x_2 + 2x_1^3x_3 + 6x_1^2x_3^2 + 9x_1^2x_2^2 - 6x_1^2x_2x_3 \\
& -14x_1x_2x_3^2 + 4x_1x_3^3 + 5x_3^4 - 7x_2^2x_3^2 + 16x_2^4
\end{aligned}
$$

Basic semialgebraic set:

$$\{x \in \mathbb{R}^n |\ f_i(x) \geq 0, h_i(x) = 0\}$$

Ex. $2x_1 + 5x_1^2x_2 - x_3 \geq 0$

$$5 - x_1^3 + 2x_1x_3 = 0$$



Ubiquitous in computational mathematics!

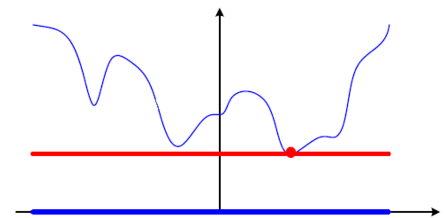
1. Polynomial optimization

$$\min_{x} p(x)$$
$$f_i(x) \leq 0$$
$$h_i(x) = 0$$

Decidable, but intractable
(includes your favorite NP-complete problem)

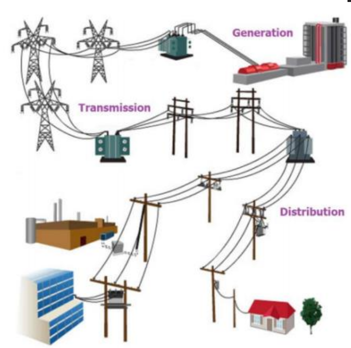
Equivalent formulation:

$$\max_{\gamma} \gamma$$
$$p(x) - \gamma \geq 0$$
$$\forall x \in \{f_i(x) \leq 0,\ h_i(x) = 0\}$$



$\}\gamma^\star$

- **Many applications:**
- Combinatorial optimization
- Option pricing with moment information
- The optimal power flow (OPF) problem
- Sensor network localization
- …



2. Infeasibility certificates in discrete optimization

- **PARTITION**

- **Input:** A list of positive integers $a_1, \dots, a_n$.

- **Question:** Can you split them into to bags such that the sum in one equals the sum in the other?

$$a = \{5,2,1,6,3,8,5,4,1,1,10\}$$

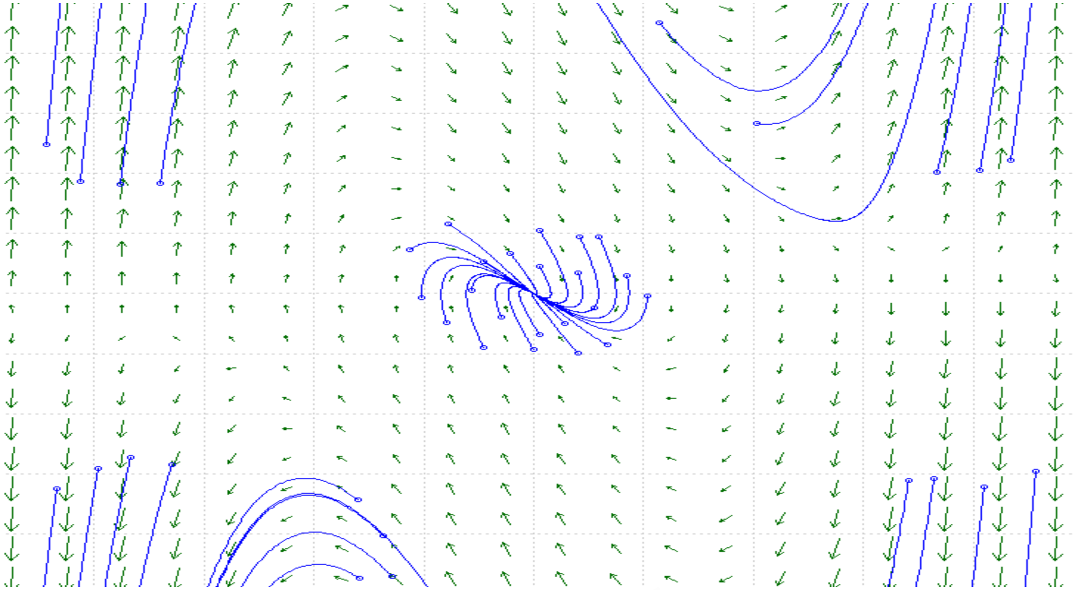
- A YES answer is easily verifiable.

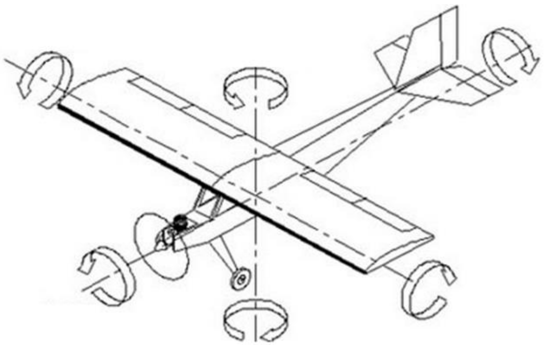
- How would you verify a NO answer?

$$p(x) = \sum_{i=1}^{n} \left(x_i^2 - 1\right)^2 + (a^T x)^2 - \epsilon \geq 0, \forall x$$

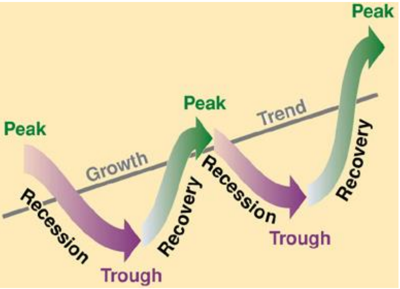
3. Stability of dynamical systems

$$\dot{x} = f(x)$$

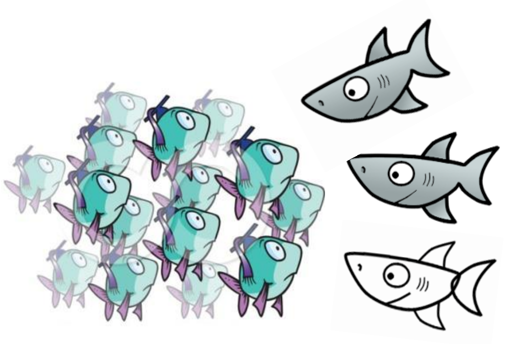




Control



Dynamics of prices



Equilibrium populations

Spread of epidemics

Robotics

Lyapunov's theorem for local stability

$$\dot{x} = f(x)$$

Existence of Lyapunov function

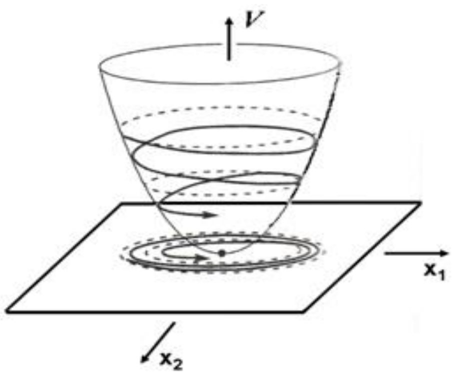
$$V(x) : \mathbb{R}^n \to \mathbb{R}$$

$$\dot{V}(x) = \langle \frac{\partial V}{\partial x}, f(x) \rangle$$

such that

$$V(x) > 0,$$

$$V(x) \le \beta \Rightarrow \dot{V}(x) < 0$$



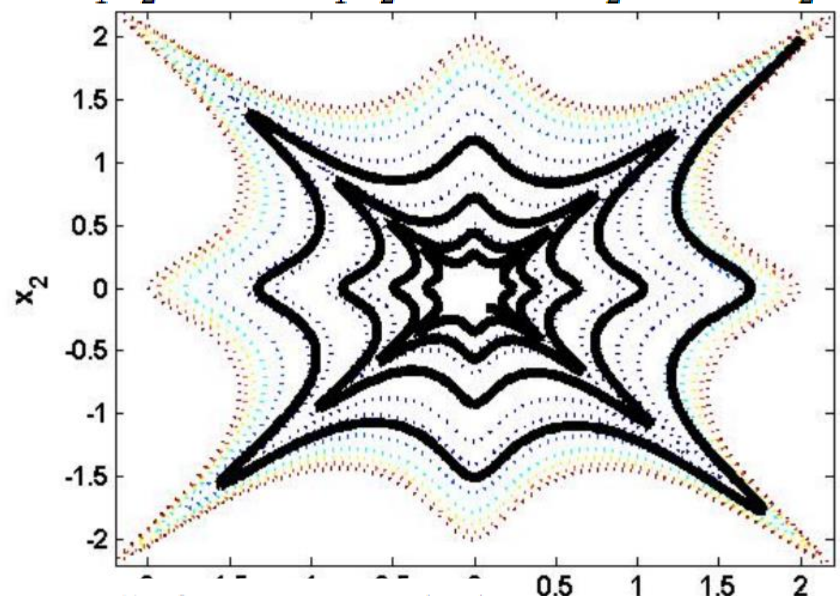
implies $\{x \mid V(x) \le \beta\}$ is in the region of attraction (ROA).

Global stability

$$\begin{aligned} V(x) &\ \text{sos} \\ -\dot{V}(x) &\ \text{sos} \end{aligned} \Rightarrow \begin{aligned} V(x) &> 0 \\ -\dot{V}(x) &> 0 \end{aligned} \Rightarrow \text{GAS}$$

Example.

$$\begin{aligned} \dot{x}_1 &= -0.15x_1^7 + 200x_1^6x_2 - 10.5x_1^5x_2^2 - 807x_1^4x_2^3 + 14x_1^3x_2^4 + 600x_1^2x_2^5 - 3.5x_1x_2^6 + 9x_2^7 \\ \dot{x}_2 &= -9x_1^7 - 3.5x_1^6x_2 - 600x_1^5x_2^2 + 14x_1^4x_2^3 + 807x_1^3x_2^4 - 10.5x_1^2x_2^5 - 200x_1x_2^6 - 0.15x_2^7 \end{aligned}$$



Output of SDP solver:

$$\begin{aligned} V = {} & 0.02x_1^8 + 0.015x_1^7x_2 + 1.743x_1^6x_2^2 - 0.106x_1^5x_2^3 - 3.517x_1^4x_2^4 \\ & +0.106x_1^3x_2^5 + 1.743x_1^2x_2^6 - 0.015x_1x_2^7 + 0.02x_2^8. \end{aligned}$$

How would you prove nonnegativity?

Ex. Decide if the following polynomial is nonnegative:

$$\begin{aligned} p(x) \;=\; & x_1^4 - 6x_1^3x_2 + 2x_1^3x_3 + 6x_1^2x_3^2 + 9x_1^2x_2^2 - 6x_1^2x_2x_3 \\ & -14x_1x_2x_3^2 + 4x_1x_3^3 + 5x_3^4 - 7x_2^2x_3^2 + 16x_2^4 \end{aligned}$$

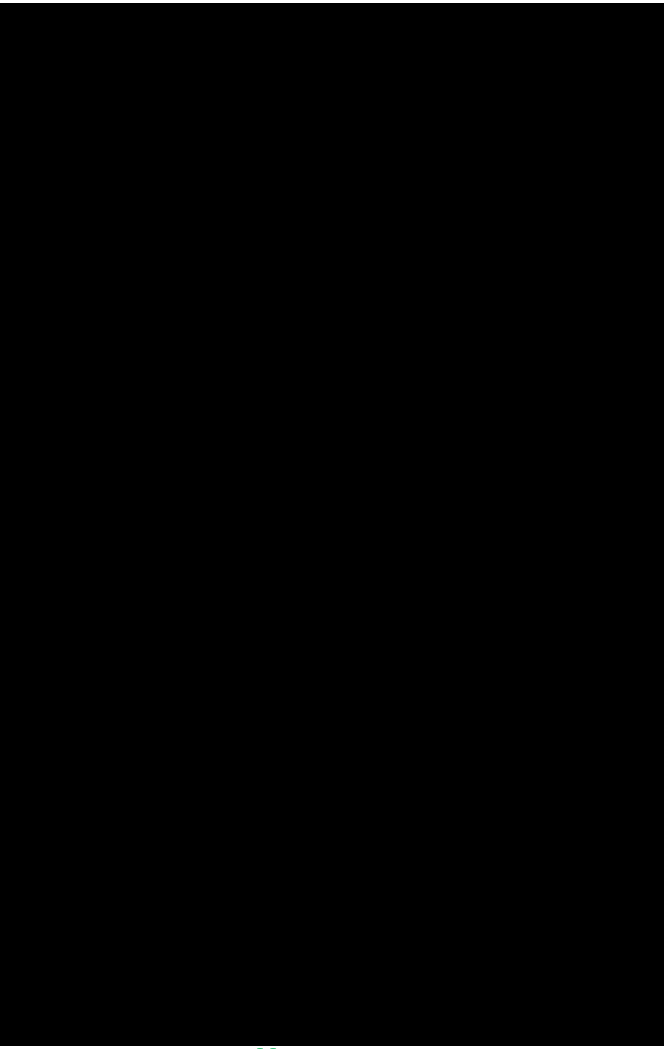
- Not so easy! (In fact, **NP-hard for degree ≥ 4**)
- But what if I told you:

$$\begin{aligned} p(x) \;=\; & (x_1^2 - 3x_1x_2 + x_1x_3 + 2x_3^2)^2 + (x_1x_3 - x_2x_3)^2 \\ & +(4x_2^2 - x_3^2)^2. \end{aligned}$$

Natural question:

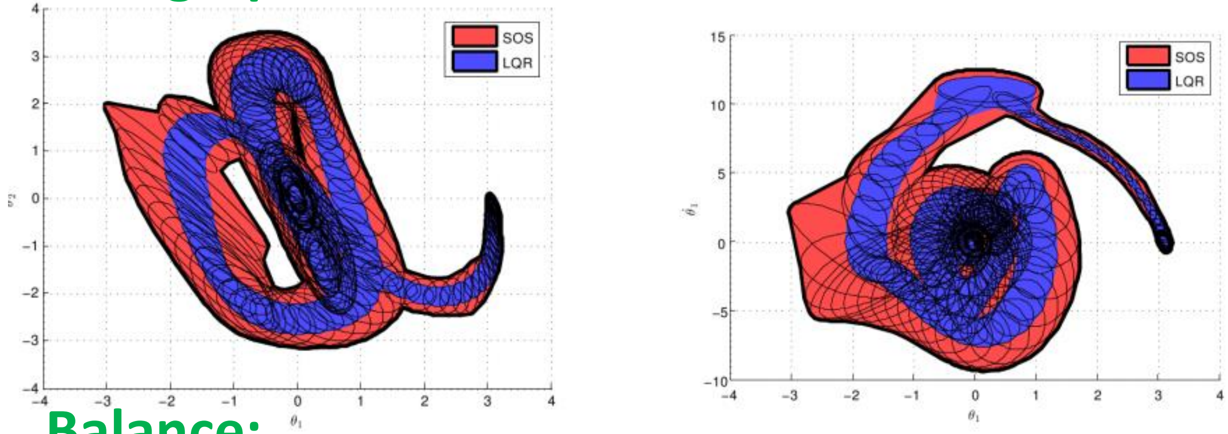
- Is it any easier to test for a sum of squares (SOS) decomposition?
  - Yes! Can be reduced to a **semidefinite program (SDP)**!
  - Can be solved to arbitrary accuracy in polynomial time.
- Extends to the "local" case (Positivstellensatz, etc.)

Local stability – SOS on the Acrobot

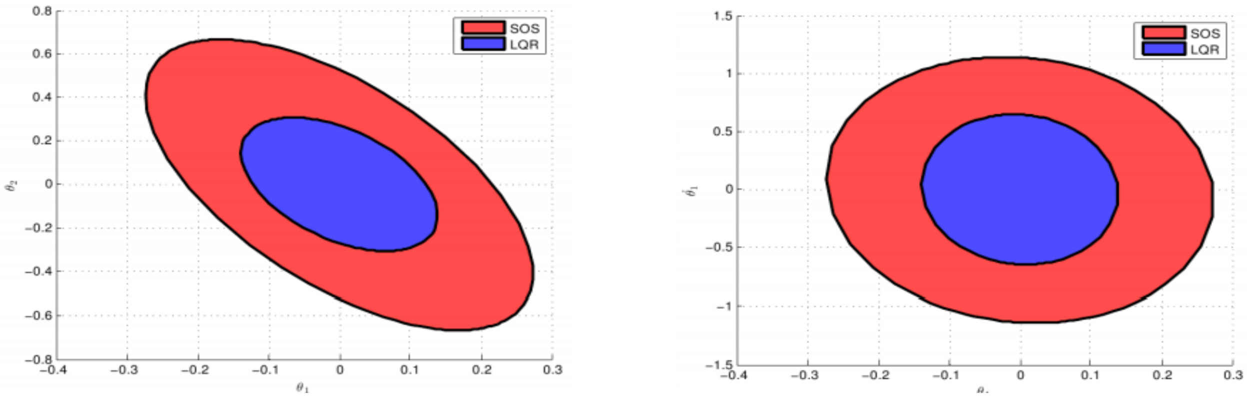


Controller
designed by SOS

Swing-up:



Balance:



[Majumdar, AAA, Tedrake]
(Best paper award - *IEEE Conf. on Robotics and Automation*, ’13)

Practical limitations of SOS

- **Scalability** is a pain in the (_|_)

Thm: ***p(x)*** of degree ***2d*** is sos if and only if

$$p(x) = z^T Q z \quad Q \succeq 0$$

$$z = [1, x_1, x_2, \ldots, x_n, x_1 x_2, \ldots, x_n^d]^T$$

- The size of the Gram matrix is:

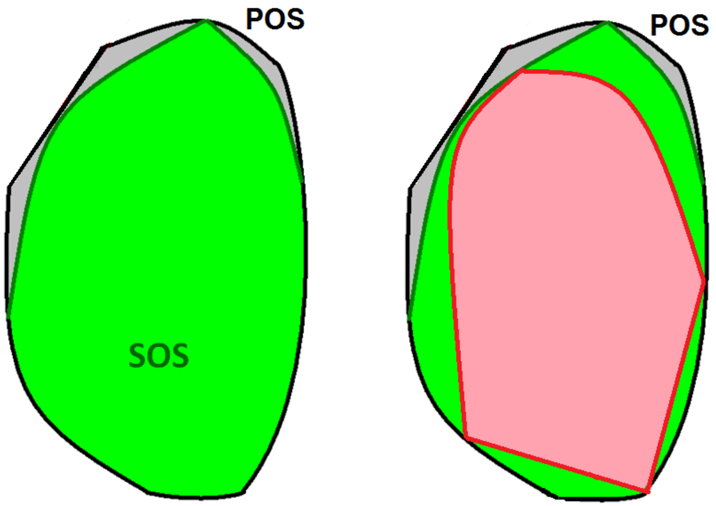
$$\binom{n+d}{d} \times \binom{n+d}{d}$$

- Polynomial in n for fixed d, but grows quickly
  - **The semidefinite constraint is expensive**
- E.g., local stability analysis of a 20-state cubic vector field is typically an SDP with ~1.2M decision variables and ~200k constraints

Many interesting approaches to tackle this issue…

- Techniques for exploiting structure (e.g., symmetry and sparsity)
    - [Gatermann, Parrilo], [Vallentin], [de Klerk, Sotirov], …
- Customized algorithms (e.g., first order or parallel methods)
    - [Bertsimas, Freund, Sun], [Nie, Wang], …

Our approach [AAA, Majumdar]:



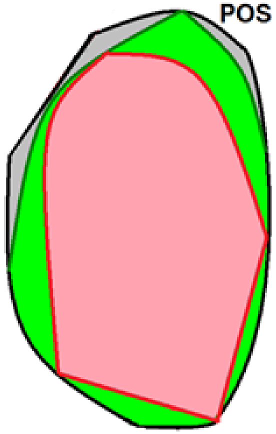
- Let's not work with SOS to begin with…
- Give other sufficient conditions for non
  perhaps stronger than SOS, but hopefully cheaper

Not totally clear a priori how to do this…

Consider, e.g., the following two sets:

1) All polynomials that are **sums of 4th powers of polynomials**
2) All polynomials that are **sums of 3 squares of polynomials**

Both sets are clearly inside the SOS cone



- But linear optimization over either set is intractable!
- So set inclusion doesn't mean anything in terms of complexity
- We have to work a bit harder…

dsos and sdsos

Defn. A polynomial ***p*** is *diagonally-dominant-sum-of-squares* ***(dsos)*** if it can be written as:

$$p = \sum_i \alpha_i m_i^2 + \sum_{i,j} \beta_{ij}^+ (m_i + m_j)^2 + \beta_{ij}^- (m_i - m_j)^2,$$

for some monomials m_i, m_j
and some nonnegative constants $\alpha_i, \beta_{i,j}$.

Defn. A polynomial ***p*** is *scaled-diagonally-dominant-sum-of-squares* ***(sdsos)*** if it can be written as:

$$p = \sum_i \alpha_i m_i^2 + \sum_{i,j} (\beta_i^+ m_i + \gamma_j^+ m_j)^2 + (\beta_i^- m_i - \gamma_j^- m_j)^2,$$

for some monomials m_i, m_j
and some constants $\alpha_i \geq 0, \beta_i, \gamma_i$.

Obvious: $DSOS_{n,d} \subseteq SDSOS_{n,d} \subseteq SOS_{n,d} \subseteq POS_{n,d}$

r-dsos and r-sdsos

Defn. A polynomial ***p*** is *r-diagonally-dominant-sum-of-squares* ***(r-dsos)*** if

$$p \cdot \left(\sum_i x_i^2\right)^r$$

is dsos.

Defn. A polynomial ***p*** is *r-scaled-diagonally-dominant-sum-of-squares* ***(r-sdsos)*** if

$$p \cdot \left(\sum_i x_i^2\right)^r$$

is sdsos.

Easy: $rDSOS_{n,d} \subseteq rSDSOS_{n,d} \subseteq POS_{n,d}, \forall r.$

$$rDSOS_{n,d} \subseteq (r+1)DSOS_{n,d}, \forall r$$

$$rSDSOS_{n,d} \subseteq (r+1)SDSOS_{n,d}, \forall r.$$

dd and sdd matrices

Defn. A symmetric matrix A is *diagonally dominant* ***(dd)*** if

$$a_{ii} \geq \sum_{j \neq i} |a_{ij}| \;\; for \;\; all \;\; i.$$

Defn*. A symmetric matrix A is *scaled diagonally dominant* ***(sdd)*** if there exists a diagonal matrix $D>0$ s.t.

DAD is dd.

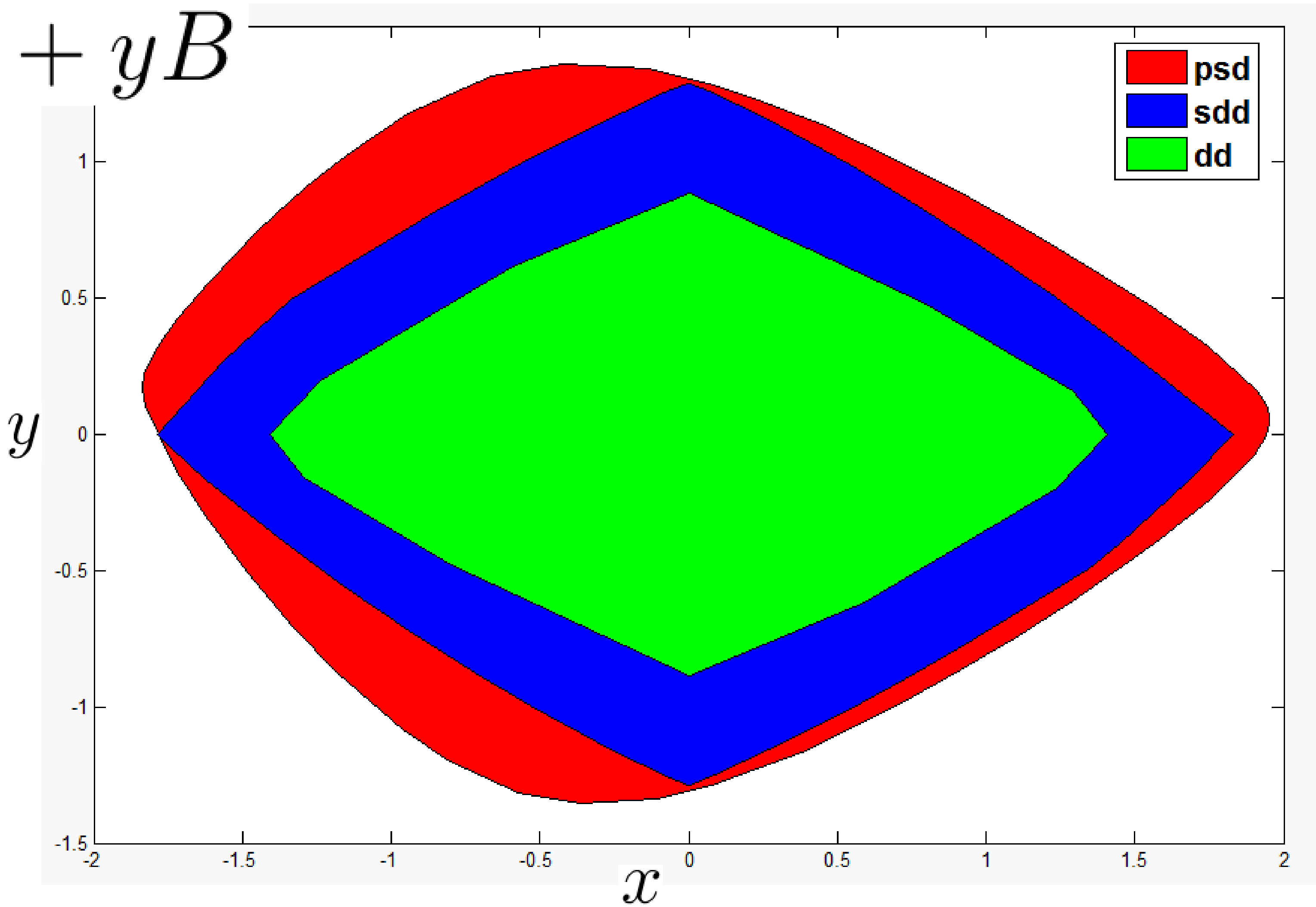
$$dd \Rightarrow sdd \Rightarrow psd$$

Greshgorin's circle theorem

*Thanks to Pablo Parrilo for telling us about sdd matrices.

$$I + xA + yB$$

A, B
10×10
random



Optimization over these sets is an SDP, SOCP, LP !!

Two natural matrix programs: DDP and SDPP

LP:

$$\min\langle C, X\rangle$$
$$A(X) = b$$
$$X \quad \text{diagonal\&nonnegative}$$

DDP:

$$\min\langle C, X\rangle$$
$$A(X) = b$$
$$X \quad \text{dd}$$

SDDP:

$$\min\langle C, X\rangle$$
$$A(X) = b$$
$$X \quad \text{sdd}$$

SDP:

$$\min\langle C, X\rangle$$
$$A(X) = b$$
$$X \succeq 0$$

From matrices to polynomials

Thm. A polynomial ***p*** is ***dsos***

$$p = \sum_i \alpha_i m_i^2 + \sum_{i,j} \beta_{ij}^{+}(m_i + m_j)^2 + \beta_{ij}^{-}(m_i - m_j)^2,$$

if and only if

$$p(x) = z^T(x) Q z(x)$$

$$Q \quad dd$$

Thm. A polynomial ***p*** is ***sdsos***

$$p = \sum_i \alpha_i m_i^2 + \sum_{i,j} (\beta_i^{+} m_i + \gamma_j^{+} m_j)^2 + (\beta_i^{-} m_i - \gamma_j^{-} m_j)^2,$$

if and only if

$$p(x) = z^T(x) Q z(x)$$

$$Q \quad sdd$$

Optimization over r-dsos and r-dsos polynomials

- Can be done by LP and SOCP respectively!
- Commercial solvers such as CPLEX and GUROBI are very mature (very fast, deal with numerical issues)
- iSOS: add-on to SPOTless (package by Megretski, Tobenkin, Permenter –MIT)

https://github.com/spot-toolbox/spotless

How well does it do?!

- We show encouraging experiments from: Control, polynomial optimization, statistics, copositive programming, cominatorial optimization, options pricing, sparse PCA, etc.
- And we'll give Positivstellensatz results (converse results)

First observation: r-dsos can outperform sos

$$p(x_1, x_2, x_3) = x_1^4 x_2^2 + x_2^4 x_3^2 + x_3^4 x_1^2 - 3x_1^2 x_2^2 x_3^2$$

psd but *not* sos!

…but it's 1-dsos.

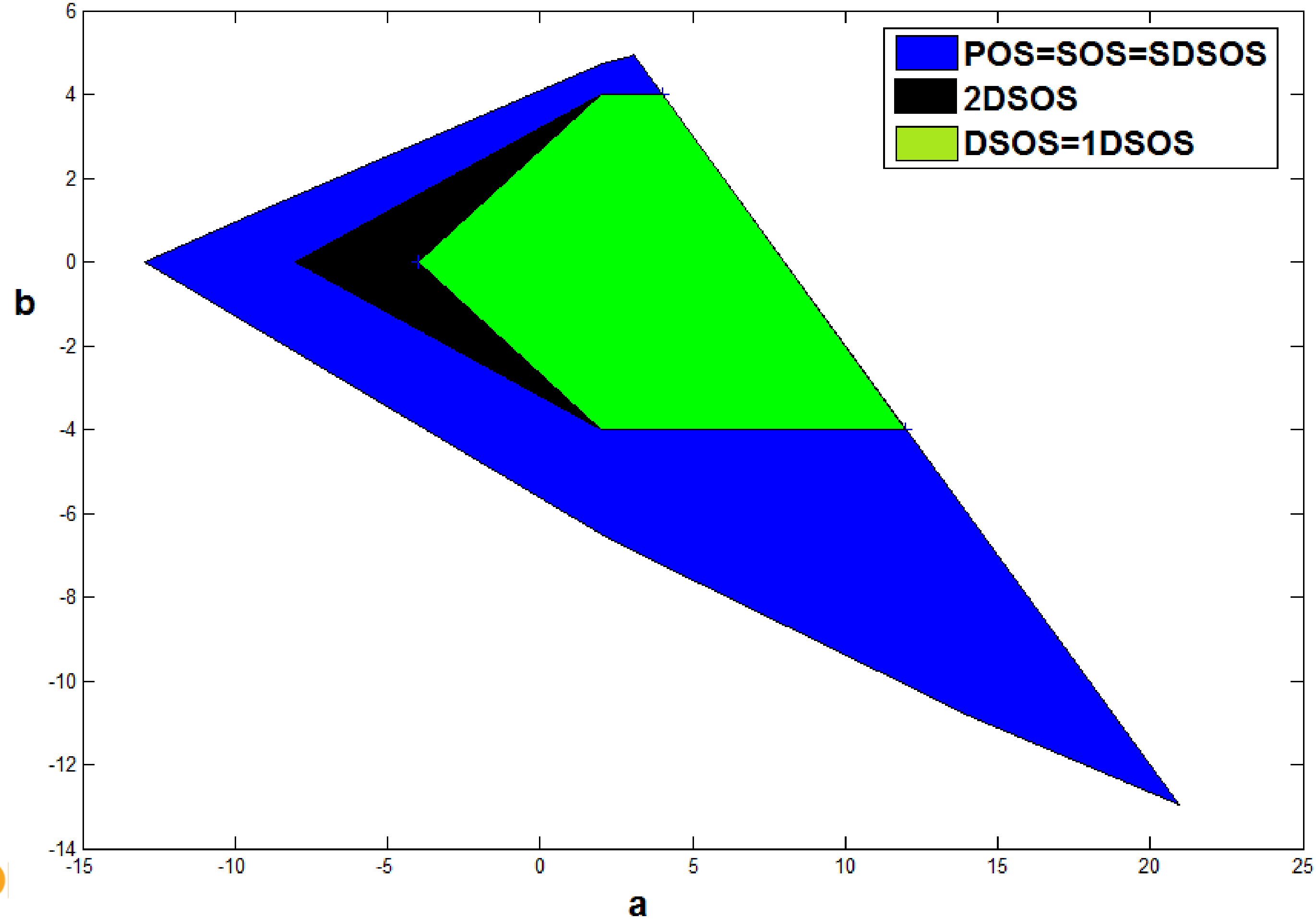
(certificate of nonnegativity using LP)

A parametric family of polynomials

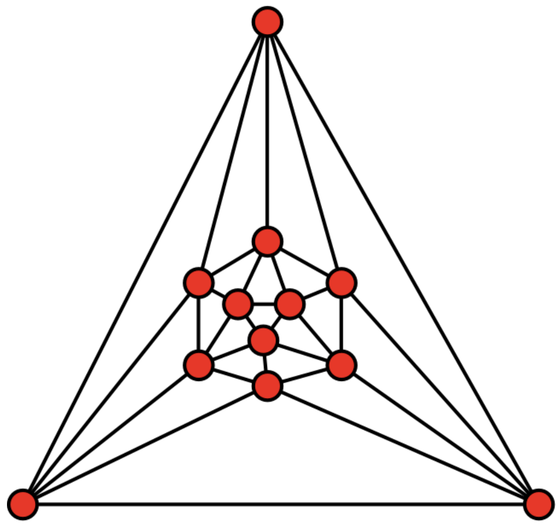
$$p(x) = 2x_1^4 + cx_2^4 + ax_1^2x_2^2 + bx_1^3x_2$$

Compactify:

$$p(x) = 2x_1^4 + (8 - a - b)x_2^4 + ax_1^2x_2^2 + bx_1^3x_2$$



Maximum Clique



Can be reformulated (via Motzkin-Straus) as a copositive program
→ positivity of a quartic

Polynomial optimization problem in 12 variables

Upper bound on max clique:

- **dsos:** 6.0000
- **sdsos:** 6.0000
- **1-dsos:** 4.3333
- **1-sdsos:** 4.3333
- **2-dsos:** 3.8049
- **2-sdsos:** 3.6964
- **sos:** 3.2362

"Polya-based LP"

- **Level 0:** ∞
- **Level 1:** ∞
- **Level 2:** 6.0000

- **r-dsos LP *guaranteed* to give exact value for r=(max clique)^2**

Minimizing a form on the sphere

$$\min_{x \in \mathcal{S}^{n-1}} p(x)$$

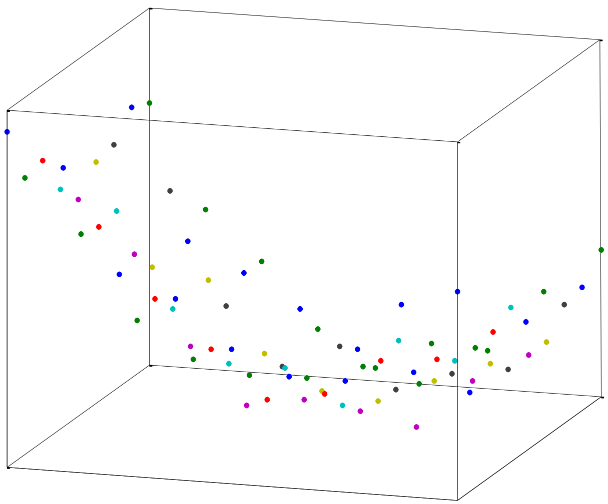
- degree=4; all coefficients present – generated randomly
- PC: 3.4 GHz, 16 Gb RAM

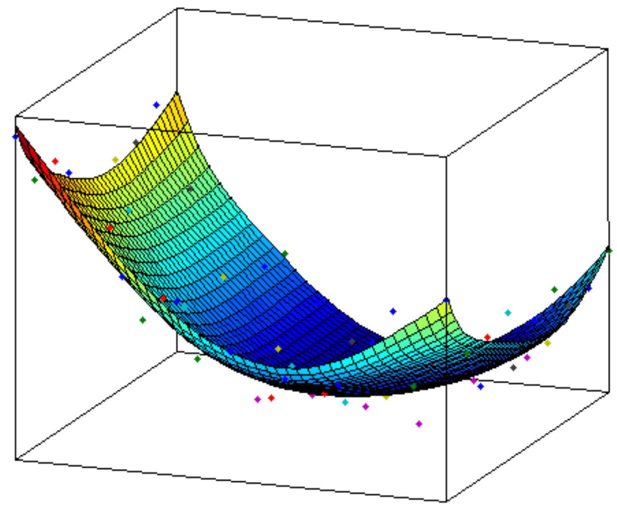
n=10	Lower bound	Run time (secs)	n=15	Lower bound	Run time (secs)	n=20	Lower bound	Run time (secs)
sos (sedumi)	-1.920	1.01	**sos (sedumi)**	-3.263	165.3	**sos (sedumi)**	-3.579	5749
sos (mosek)	-1.920	0.184	**sos (mosek)**	-3.263	5.537	**sos (mosek)**	-3.579	79.06
sdsos	-5.046	0.152	**sdsos**	-10.433	0.444	**sdsos**	-17.333	1.935
dsos	-5.312	0.067	**dsos**	-10.957	0.370	**dsos**	-18.015	1.301
BARON	-175.4	0.35	**BARON**	-1079.9	0.62	**BARON**	-5287.9	3.69

n=30	Lower bound	Run time (secs)	n=40	Lower bound	Run time (secs)	n=50	Lower bound	Run time (secs)
sos (sedumi)	---------	∞	**sos (sedumi)**	---------	∞	**sos (sedumi)**	---------	∞
sos (mosek)	---------	∞	**sos (mosek)**	---------	∞	**sos (mosek)**	---------	∞
sdsos	-36.038	9.431	**sdsos**	-61.248	53.95	**sdsos**	-93.22	100.5
dsos	-36.850	8.256	**dsos**	-62.2954	26.02	**dsos**	-94.25	72.79
BARON	-28546.1							

Convex regression

300 points in $\mathbb{R}^{20}$





Observation: $e^{||x||} + \text{ noise}$

Best convex polynomial fit of degree d

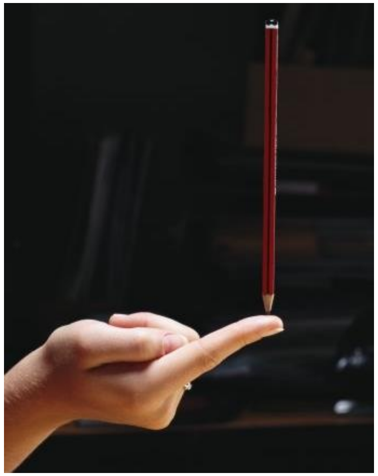
(sd)sos constraint in 40 variables:

$$y^T H(x) y \text{ (sd)sos}$$

d=2	Max Error	Run time (secs)
sos (mosek)	21.282	~1
sdsos	33.918	~1
dsos	35.108	~1

d=4	Max Error	Run time (secs)
sos (mosek)	---------	∞
sdsos	12.936	231
dsos	14.859	150

Stabilizing the inverted N-link pendulum (2N states)

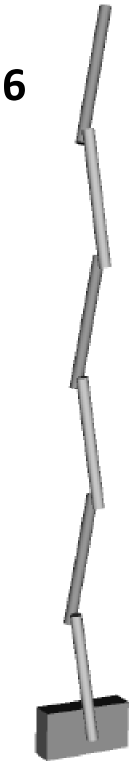


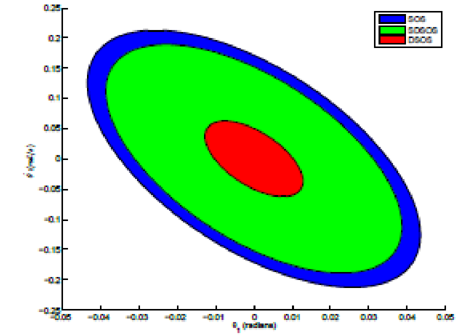
N=1

N=2

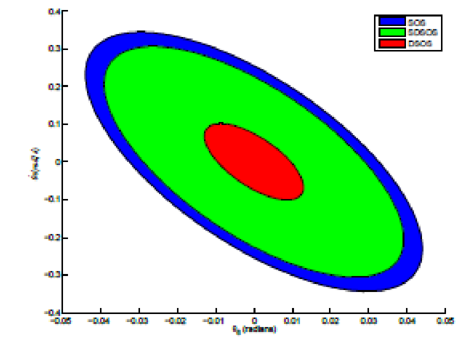
https://www.youtube.com/watch?v=FeCwtvrD76I

N=6





(a) θ_1-$\dot{\theta}_1$ subspace.



(b) θ_6-$\dot{\theta}_6$ subspace.

Runtime:

2N (# states)	4	6	8	10	12	14	16	18	20	22
DSOS	< 1	0.44	2.04	3.08	9.67	25.1	74.2	200.5	492.0	823.2
SDSOS	< 1	0.72	6.72	7.78	25.9	92.4	189.0	424.74	846.9	1275.6
SOS (SeDuMi)	< 1	3.97	156.9	1697.5	23676.5	∞	∞	∞	∞	∞
SOS (MOSEK)	< 1	0.84	16.2	149.1	1526.5	∞	∞	∞	∞	∞

ROA volume ratio:

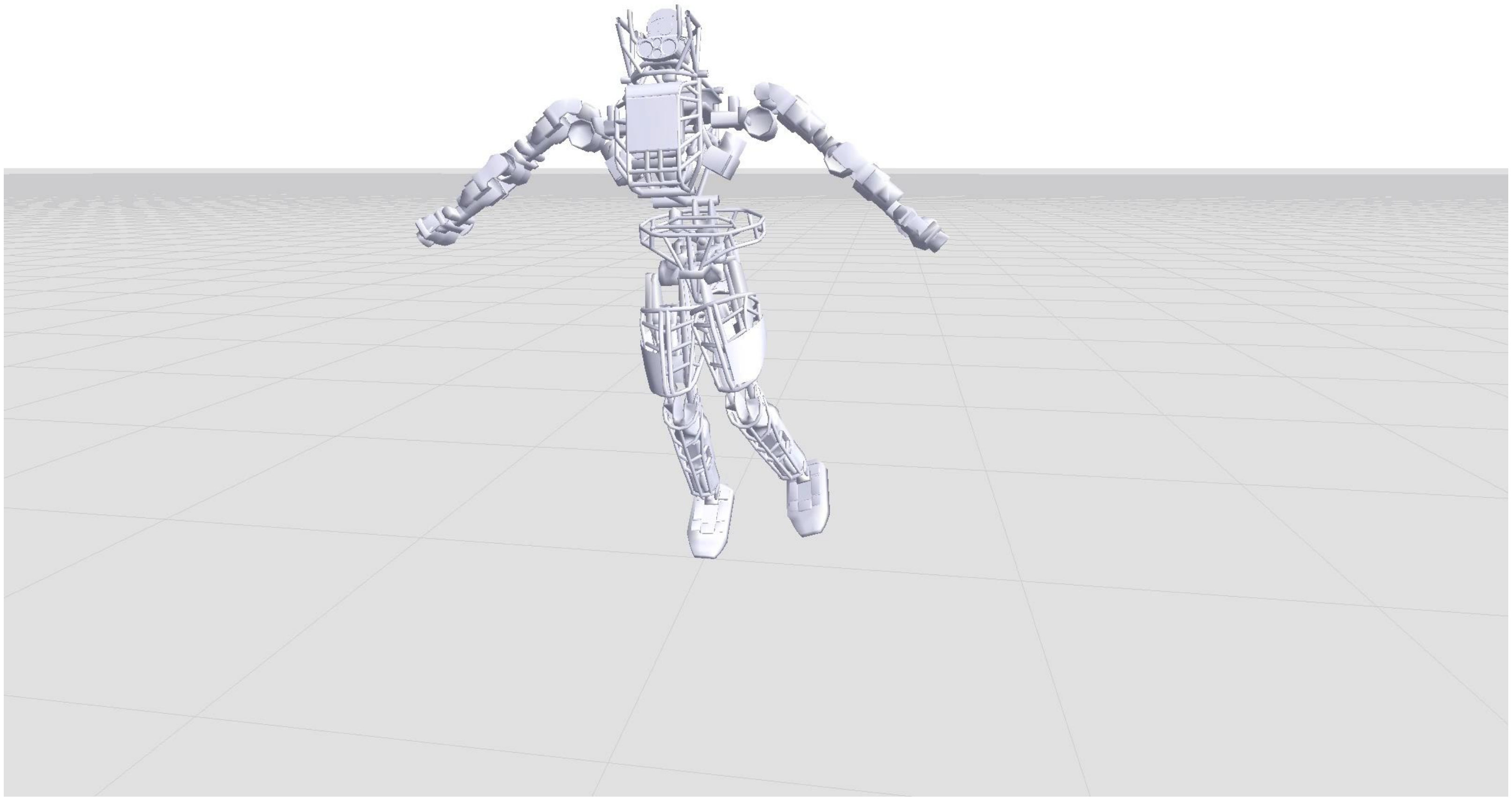
2N (states)	4	6	8	10	12
ρ_{dsos}/ρ_{sos}	0.38	0.45	0.13	0.12	0.09
ρ_{sdsos}/ρ_{sos}	0.88	0.84	0.81	0.79	0.79

(w/ Majumdar, Tedrake)

Stabilizing ATLAS

- 30 states 14 control inputs Cubic dynamics

(way beyond reach of SOS techniques)



Stabilizing controller designed by **SDSOS Optimization**

PRINCETON UNIVERSITY ORFE

[Majumdar, AAA, Tedrake]

Converse results 1&2

Thm. Any **even** positive definite form p **is r-dsos** for some r.

- Hence proof of positivity can always be found with LP
- Proof follows from a result of **Polya (1928)** on Hilbert's 17th problem
- Even forms include, e.g., **copositive programming**!

Thm. Any positive definite **bivariate** form p **is r-sdsos** for some r.

- Proof follows from a result of **Reznick (1995)**
  - $p.||x||^r$ will always become a sum of powers of linear forms for sufficiently large r.

PRINCETON UNIVERSITY ORFE

Converse result 3 & Polynomial Optimization

Thm. For any positive definite form $\boldsymbol{p}$, there exists an integer r and a **polynomial $\boldsymbol{q}$ of degree $\boldsymbol{r}$** such that

$\boldsymbol{q}$ is dsos and **$\boldsymbol{pq}$ is dsos.**

- Search for q is an LP
- Such a q is a certificate of nonnegativity of p
- Proof follows from a result of **Habicht (1940)** on Hilbert's 17th problem

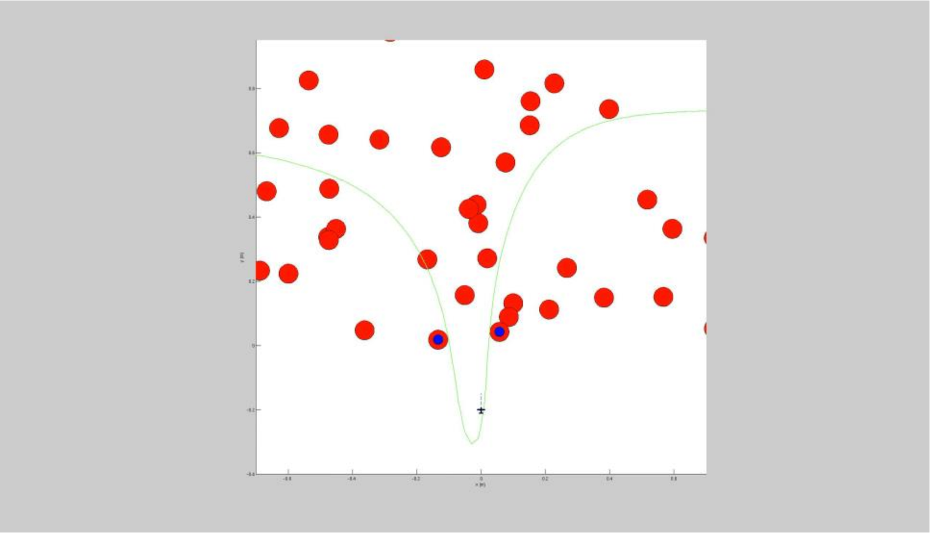
$$\min_x p(x)$$
$$f_i(x) \leq 0$$
$$h_i(x) = 0$$

- Similar to the **Lasserre/Parrilo SDP hierarchies**, **polynomial optimization can be solved to global optimality using hierarchies of LP and SOCP coming from dsos and sdsos.**

Ongoing directions…

Move towards
real-time algebraic optimization

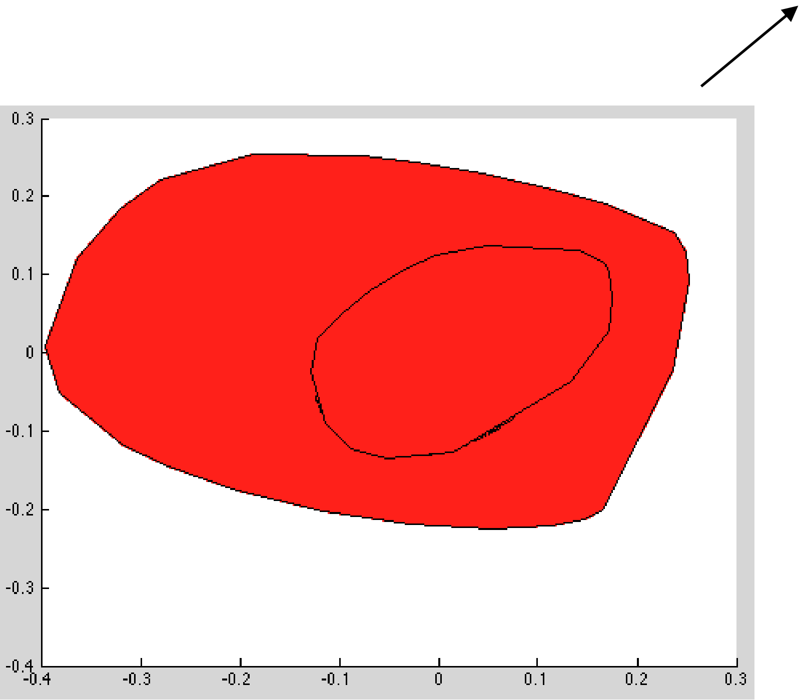
- e.g., barrier certificates
[Prajna, Jadbabaie, Pappas]



(w/ A. Majumdar, MIT)

Iterative DSOS via
- Column generation

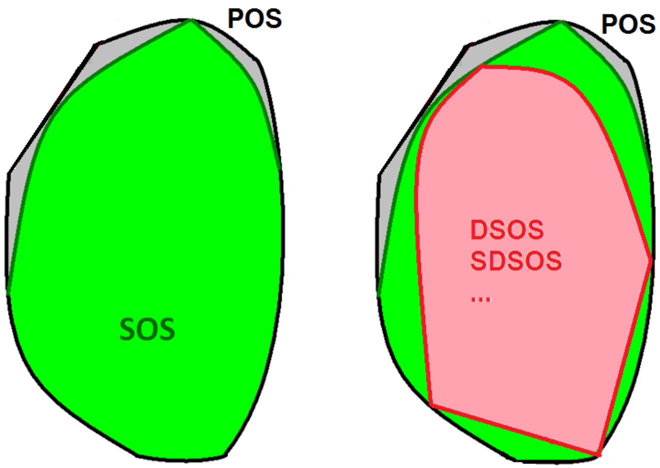
- Cholesky change of basis



(w/ S. Dash, IBM)

Main messages...

- Inner approximations to SOS cone
  Move away from SDP towards LP and SOCP
- Orders of magnitude more scalable
  - Largest we have solved: degree-4 in **70 variables**.
- Many theoretical guarantees still go through!
- This can be used *anywhere* SOS is used!



Want to know more? aaa.princeton.edu
mit.edu/~anirudha/