DSOS, SDSOS Optimization: More Tractable Alternatives to SOS Optimization

Amir Ali Ahmadi

Princeton University

Dept. of Operations Research and Financial Engineering (ORFE)

Joint work with:

Anirudha Majumdar (Stanford→Princeton)

CDC 2016, Las Vegas

Workshop on "Solving large-scale SDPs with applications to control, machine learning, and robotics"

PRINCETON UNIVERSITY ORFE

Optimization over nonnegative polynomials

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

Ex. Decide if the following polynomial is nonnegative:

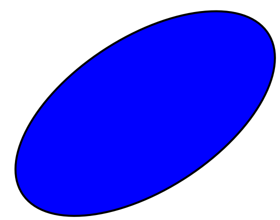
$$\begin{aligned} p(x) \;=\; & x_1^4 - 6x_1^3x_2 + 2x_1^3x_3 + 6x_1^2x_3^2 + 9x_1^2x_2^2 - 6x_1^2x_2x_3 \\ & -14x_1x_2x_3^2 + 4x_1x_3^3 + 5x_3^4 - 7x_2^2x_3^2 + 16x_2^4 \end{aligned}$$

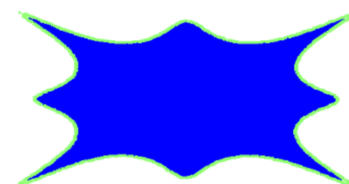
Basic semialgebraic set:

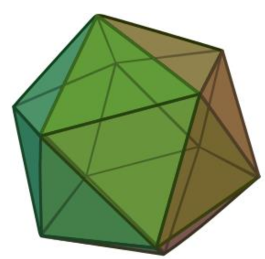
$$\{x \in \mathbb{R}^n \mid f_i(x) \geq 0, h_i(x) = 0\}$$

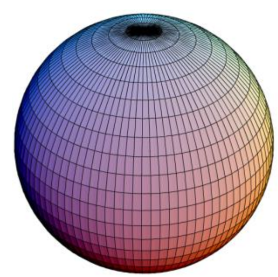
Ex. $2x_1 + 5x_1^2x_2 - x_3 \geq 0$

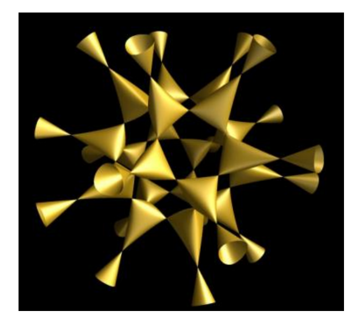
$$5 - x_1^3 + 2x_1x_3 = 0$$









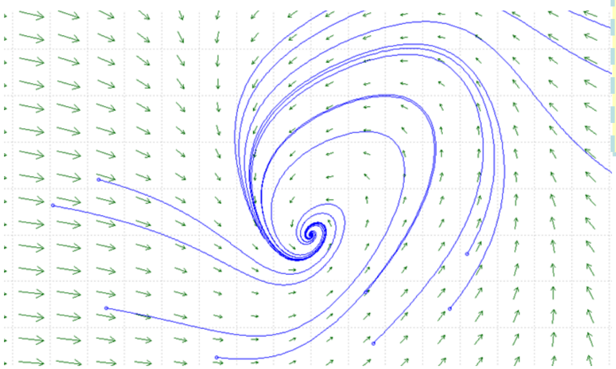


PRINCETON UNIVERSITY ORFE

Ubiquitous in computational mathematics!

Application 1: verification of dynamical systems

$$\dot{x} = f(x)$$
$$x_{k+1} = f(x_k)$$



Properties of interest:

- Stability of equilibrium points
- Boundedness of trajectories
- Invariance of sets
- Collision avoidance
- …

Lyapunov Theory

Search for functions satisfying certain **nonnegativity constraints**

E.g. - existence of a Lyapunov function

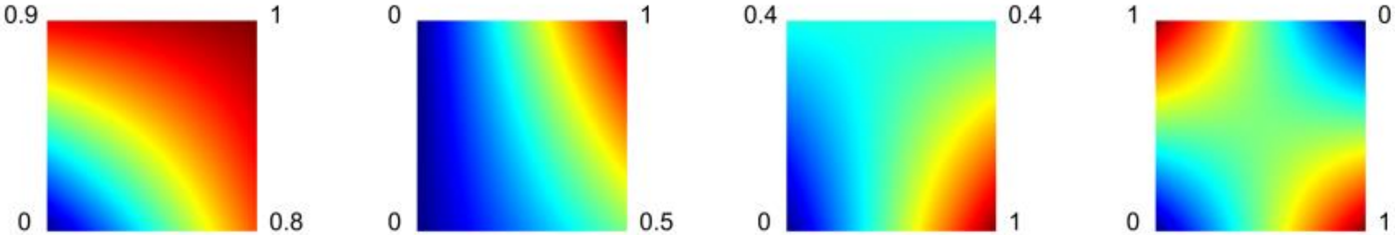
$$V(x) > 0,$$
$$V(x) \leq \beta \Rightarrow \dot{V}(x) < 0$$

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

PRINCETON UNIVERSITY ORFE

2: Statistics and Machine Learning

- **Shape-constrained regression; e.g., *monotone regression***



Monotonic (a) | Monotonic (b) | Not Monotonic (c) | Not Monotonic (d)

(From [Gupta et al.,'15])

- How to parameterize a polynomial $p(x_1, x_2)$ to enforce monotonicity over $[0,1]^2$?

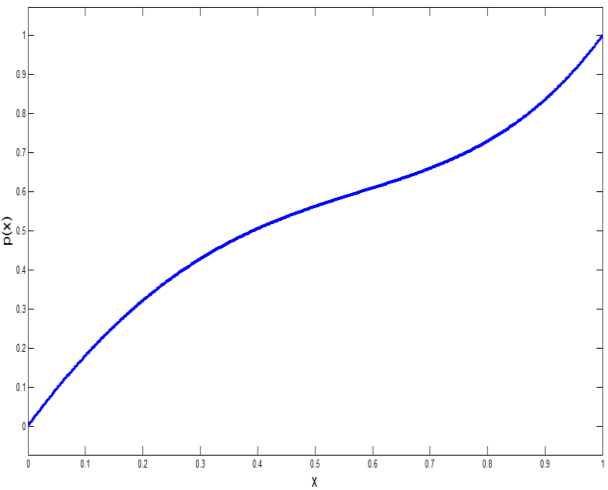
- Need its partial derivatives to be nonnegative over $[0,1]^2$.
- Let's see a simple example in one variable…

PRINCETON UNIVERSITY ORFE

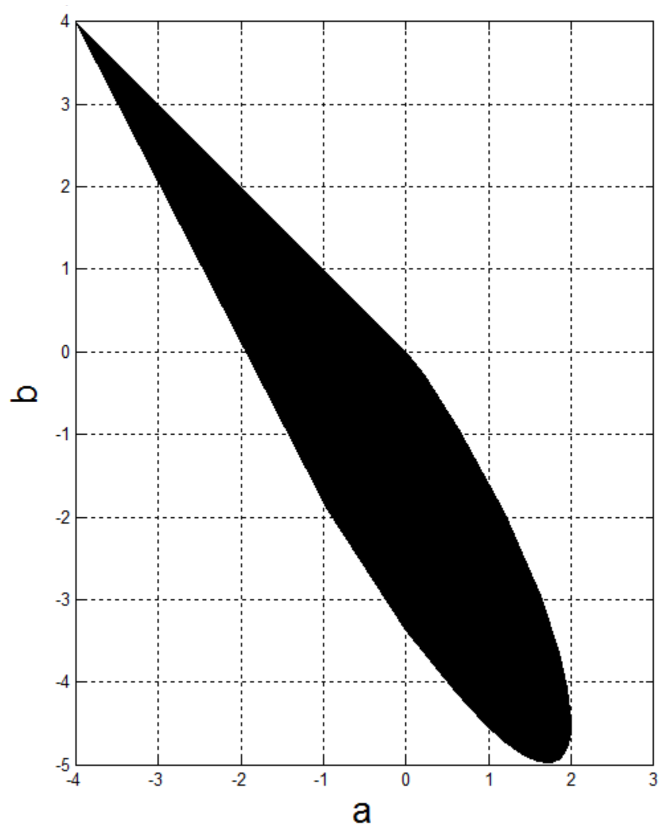
Imposing monotonicity

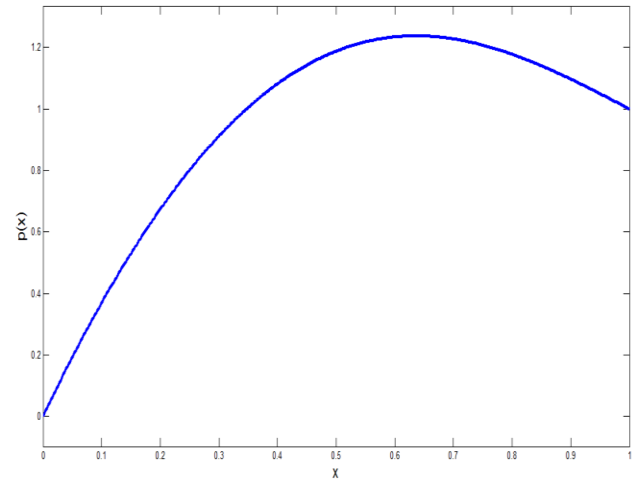
- For what values of a, b is the following polynomial monotone over $[0,1]$?

$$p(x) = x^4 + ax^3 + bx^2 - (a+b)x$$



$a = 0, b = -2$





$a = -1, b = -3$

How to prove nonnegativity?

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

Nonnegative

$\Uparrow$

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

SOS

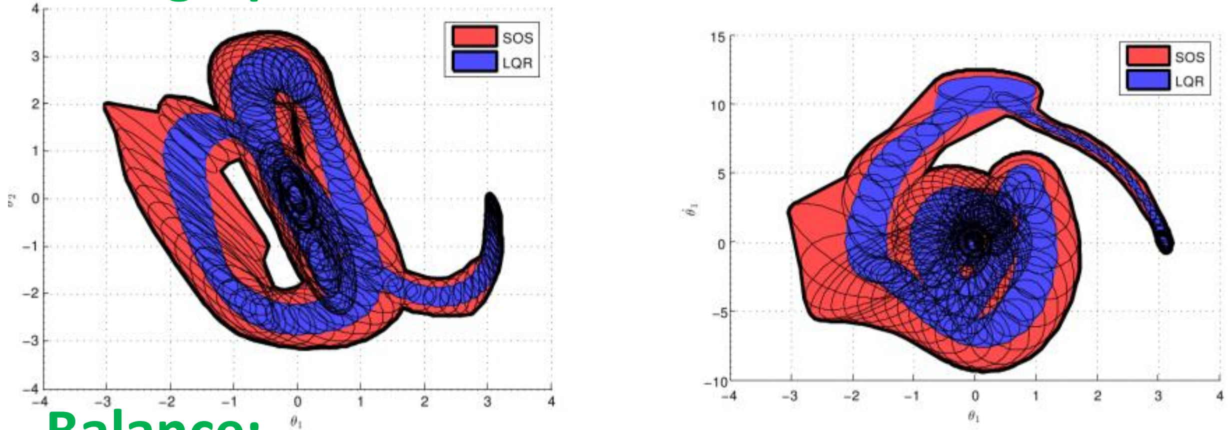
- **Extends to the constrained case:**

$$\boldsymbol{p(x) = \sigma_0(x) + \sum \sigma_i(x) g_i(x), \sigma_i(x)\ SOS \Rightarrow p(x) \geq 0}\ \textbf{ on } \{\boldsymbol{x} |\ \boldsymbol{g_i(x) \geq 0}\}$$

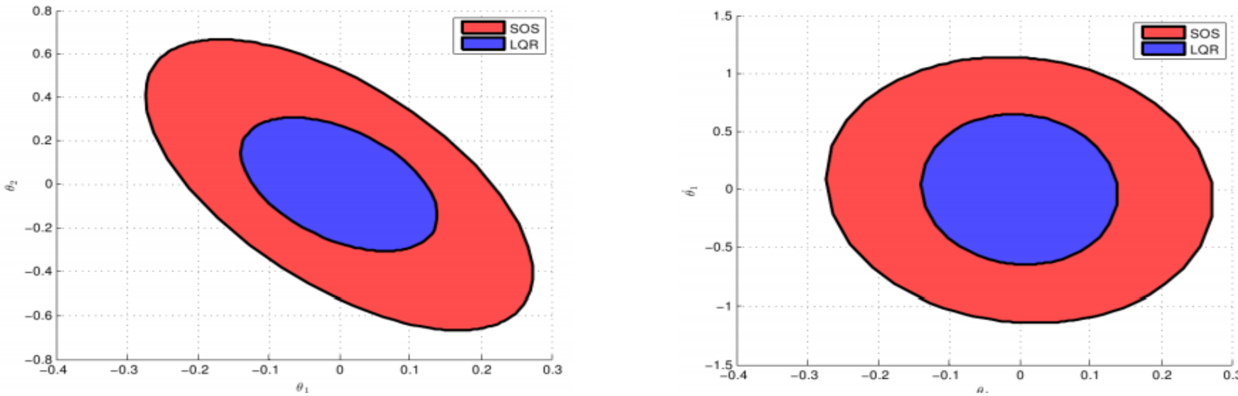
- The search for such algebraic certificates ----> SDP!!
- Can produce a hierarchy; connections to Hilbert's 17^{th} problem, etc.
- Fundamental work of many: [Lasserre, Nesterov, Parrilo, Shor, …]

Local stability – SOS on the Acrobot

Swing-up:



Balance:



Controller designed by SOS

[Majumdar, AAA, Tedrake]

(Best paper award - *IEEE Conf. on Robotics and Automation*)

Practical limitations of SOS

- **Scalability** is a nontrivial challenge!

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

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

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

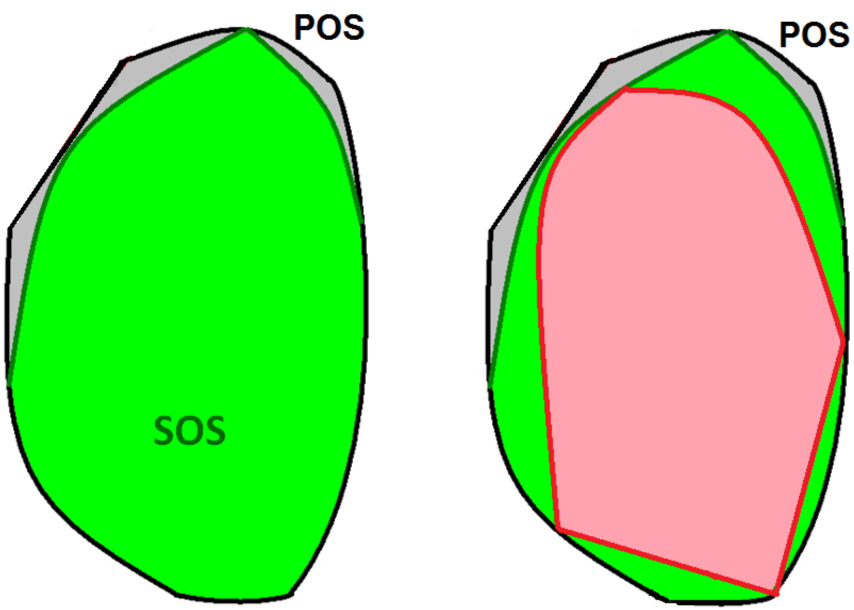
- The size of the Gram matrix is:

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

- 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

Simple idea…

- Let’s not work with SOS…
- Give other sufficient conditions for nonnegativity that are **perhaps stronger than SOS, but hopefully cheaper**



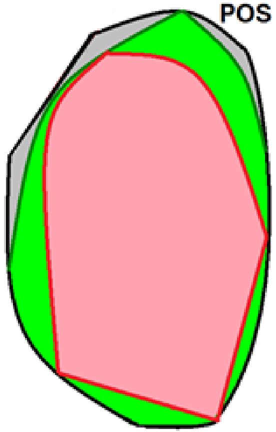
[AAA, Majumdar]

Not any set inside SOS would work…

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:

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

for some monomials m_i, m_j
and nonnegative scalars $\alpha_i, \beta_{ij}^+, \beta_{ij}^-$

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

$$p(x) = \sum_i \alpha_i m_i^2(x) + \sum_{i,j} (\hat{\beta}_{ij}^+ m_i(x) + \tilde{\beta}_{ij}^+ m_j(x))^2 + \sum_{ij} (\hat{\beta}_{ij}^- m_i(x) - \tilde{\beta}_{ij}^- m_j(x))^2,$$

for some monomials m_i, m_j
and scalars $\alpha_i, \hat{\beta}_{ij}^+, \tilde{\beta}_{ij}^+, \hat{\beta}_{ij}^-, \tilde{\beta}_{ij}^-$ with $\alpha_i \geq 0.$

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

r-dsos and r-sdsos

Defn. A polynomial $\boldsymbol{p}$ 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 $\boldsymbol{p}$ is *r-scaled-diagonally-dominant-sum-of-squares* ***(r-sdsos)*** if

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

is sdsos.

Allows us to develop a ***hierarchy*** of relaxations…

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.

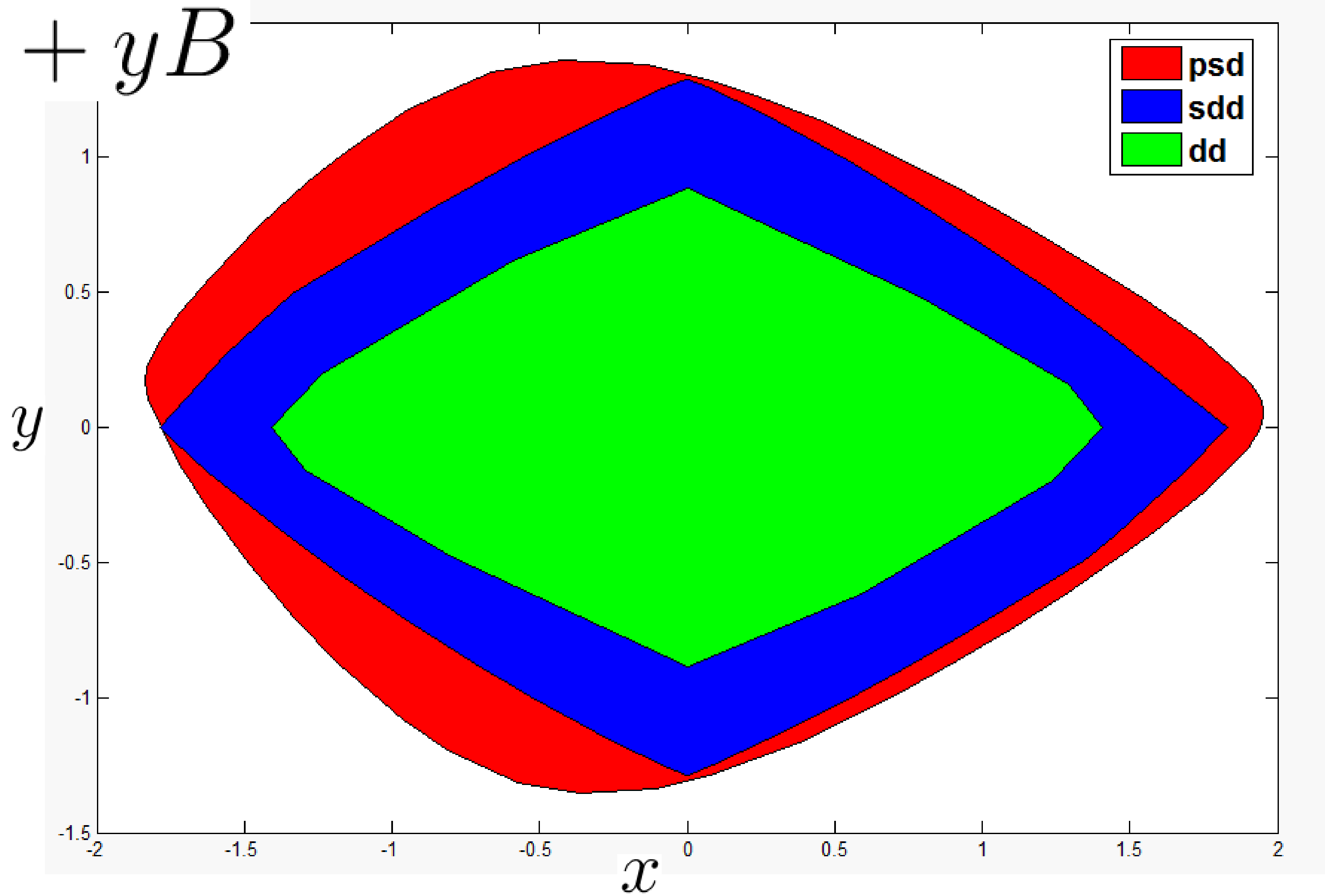
$$dd \Rightarrow sdd \Rightarrow psd$$

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!
- iSOS: add-on to SPOTless (package by Megretski, Tobenkin, Permenter –MIT)

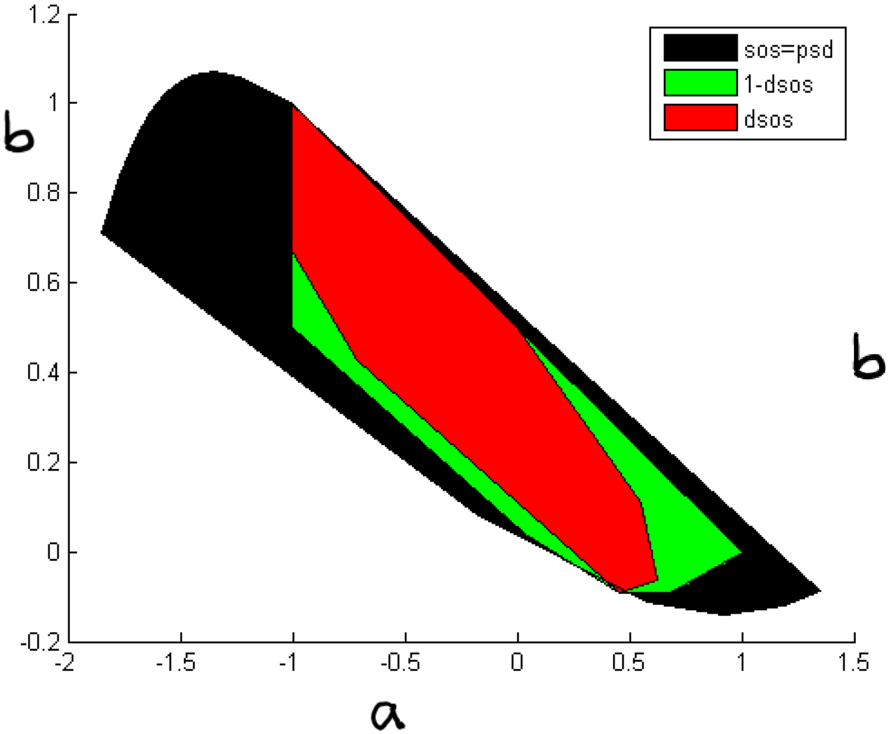
https://github.com/spot-toolbox/spotless

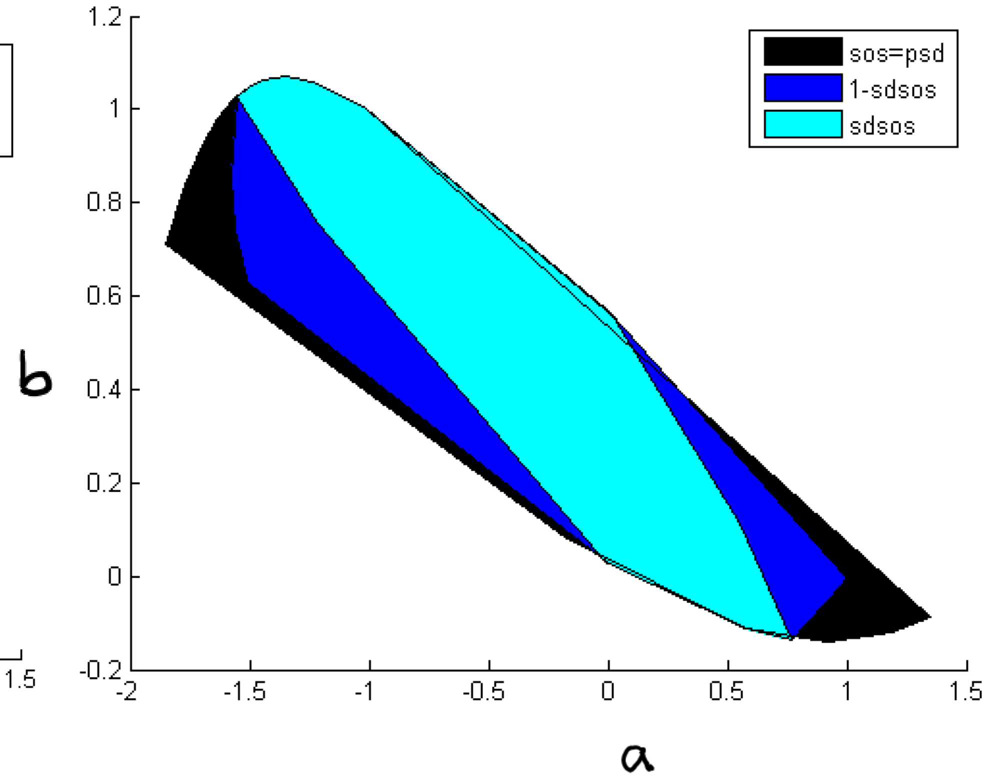
How well does it do?!

- Our paper shows encouraging experiments from: *Control, polynomial optimization, statistics, combinatorial optimization, options pricing, sparse PCA, etc.*
- And we'll give converse results

A parametric family

$$P(x)=\frac{1}{2}x_1^4+\frac{1}{2}x_2^4+a\,x_1^3x_2+b\,x_1^2x_2^2+(1-2a-4b)\,x_1x_2^3$$





Converse results

Thm. Any **even** positive definite form $\boldsymbol{p}$ **is r-dsos** for some $\boldsymbol{r}$**.**

- 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 copositive programming, nonnegative switched systems, etc.)

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)**

Converse results: stability of switched linear systems

Problem:

Given a set of $n \times n$ matrices $M = \{A_1, \dots, A_m\}$

When is the system $x_{k+1} = A_{\sigma(k)} x_k$ stable?

Joint spectral radius (JSR) of $\boldsymbol{M = \{A_1, \dots, A_m\}}$:

$$\rho(A_1, \dots, A_m) = \lim_{k \to \infty} \max_{\sigma \in \{1,\dots,m\}^k} \left|\left| A_{\sigma_k} \dots A_{\sigma_2} A_{\sigma_1} \right|\right|^{1/k}$$

Theorem:

Switched linear system is stable $\Leftrightarrow \rho(A_1, \dots, A_m) < 1$

Goal: compute upperbounds on JSR

Converse results: stability of switched linear systems

Link to polynomial nonnegativity:

$$\rho(A_1, \dots, A_m) < 1$$

$$\Leftrightarrow$$

$\exists$ a pd polynomial Lyapunov function $V(x)$ such that $V(x) - V(A_i x) > 0, \forall x \neq 0.$

Semidefinite relaxation [Parrilo, Jadbabaie]:

$$\rho(A_1, \dots, A_m) < 1$$

$$\Leftrightarrow$$

$\exists$ an sos polynomial Lyapunov function $V(x)$ such that $V(x) - V(A_i x)$ sos.

Converse results: stability of switched linear systems

Theorem (AAA,Hall): For nonnegative $\{\mathrm{A}_1, \dots, \mathrm{A}_\mathrm{m}\}$, $\rho(A_1, \dots, A_m) < 1 \Leftrightarrow$ $\exists\, r \in \mathbb{N}$ and a polynomial Lyapunov function $V(x)$ such that

$$V(x.^2) \text{ r-dsos and } V(x.^2) - V(A_i x.^2) \text{ r-dsos.} \qquad (\star)$$

Proof:

$(\Leftarrow)$ $(\star) \Rightarrow V(x) \geq 0$ and $V(x) - V(A_i x) \geq 0$ for any $x \geq 0$.

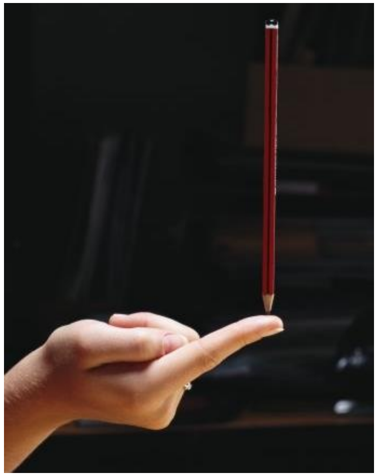
Combined to $A_i \geq 0$, this implies that $x_{k+1} = A_{\sigma(k)} x_k$ is stable for $x_0 \geq 0$.

This can be extended to any x_0 by noting that $x_0 = x_0^+ - x_0^-, x_0^+, x_0^- \geq 0$.

$(\Rightarrow)$ From theorem of Parrilo-Jadbabie, and using Polya's result as $V(x.^2)$ and $V(x.^2) - V(A_i x.^2)$ are even forms.

Larger-scale applications in control

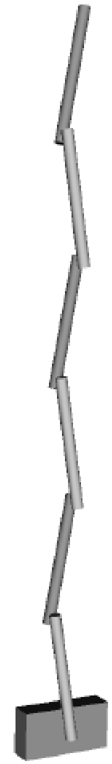
Stabilizing the inverted N-link pendulum (2N states)

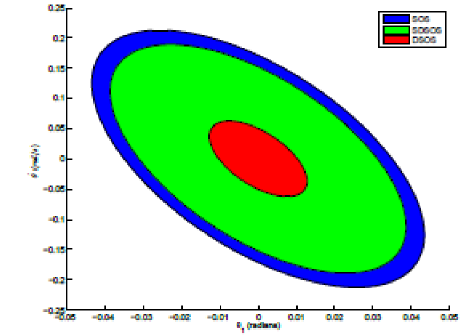


N=1

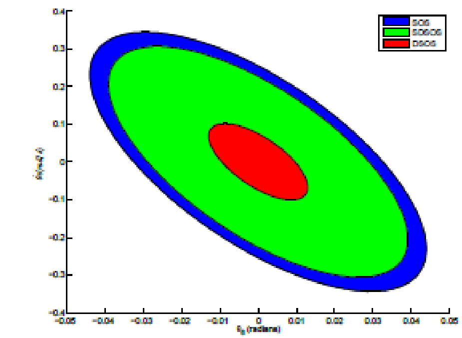
N=2

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	∞	∞	∞	∞	∞

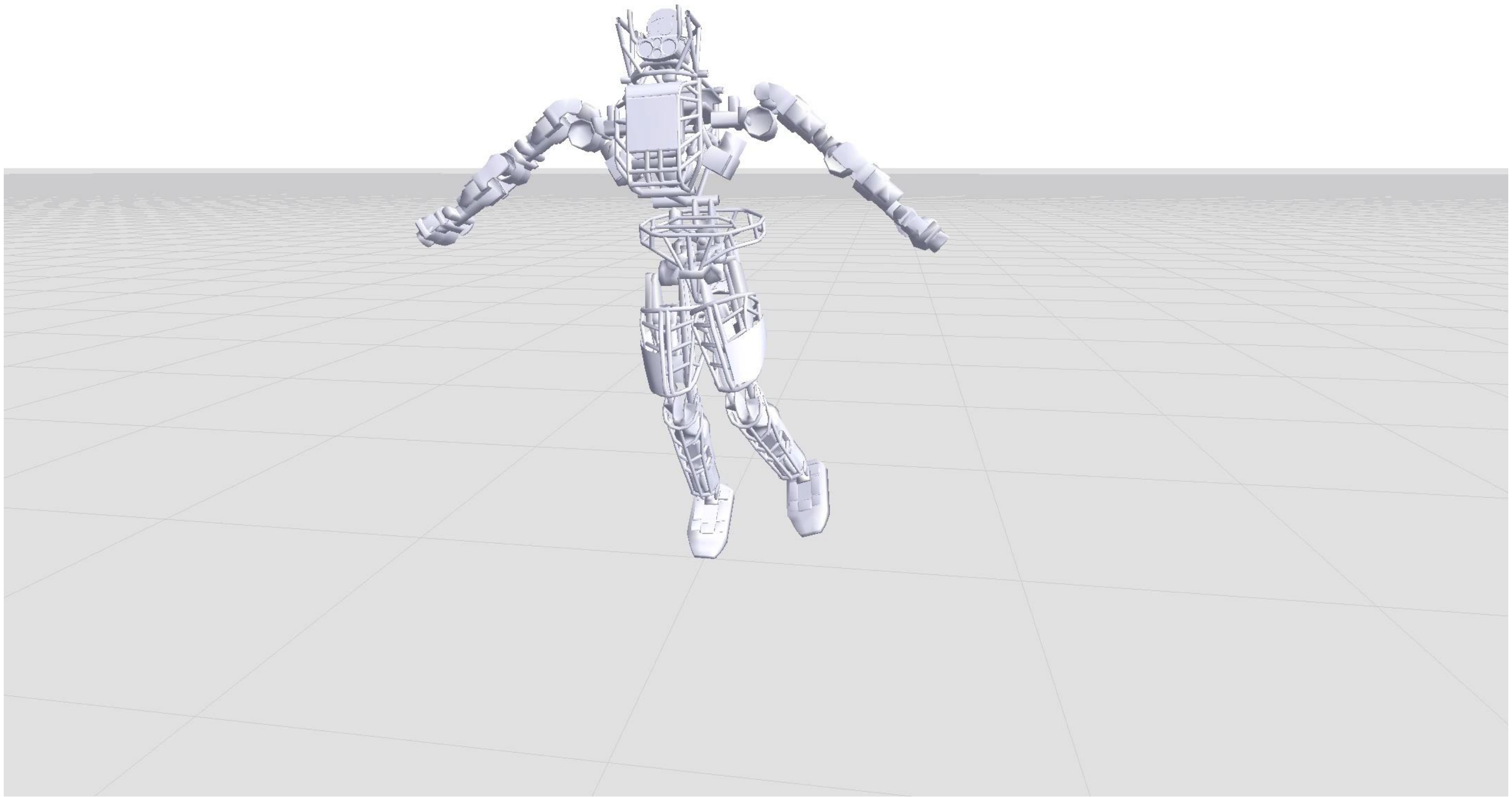
ROA volume ratio:

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



Done by **SDSOS Optimization**

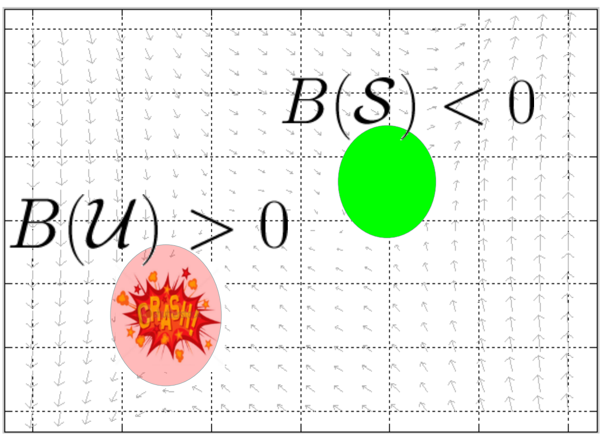
[Majumdar, AAA, Tedrake]

More recent directions...

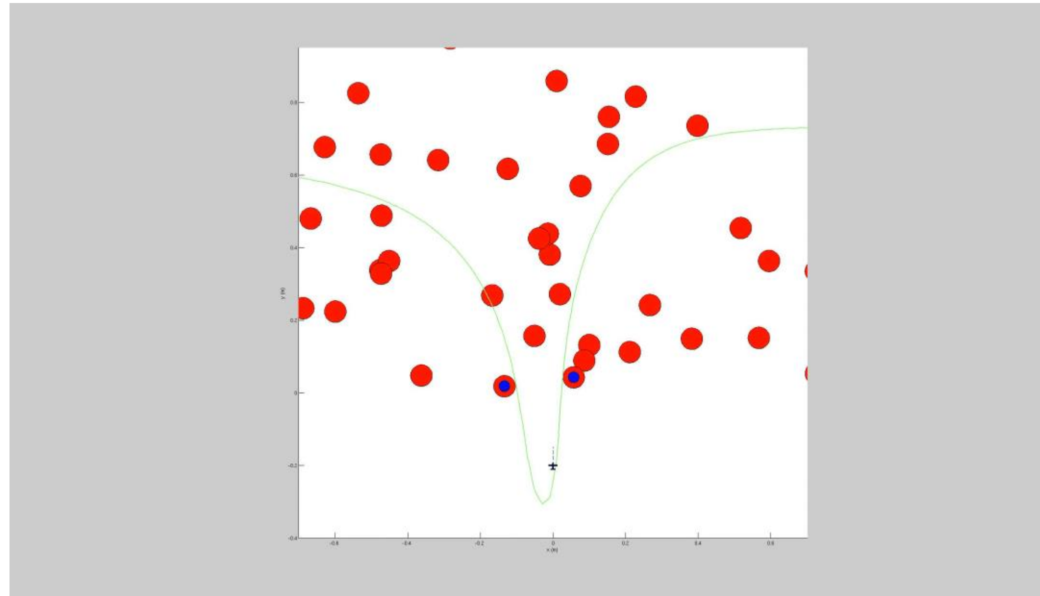
Move towards
real-time algebraic optimization

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

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



$$\dot{B} = \langle \nabla B(x), f(x) \rangle \leq 0$$



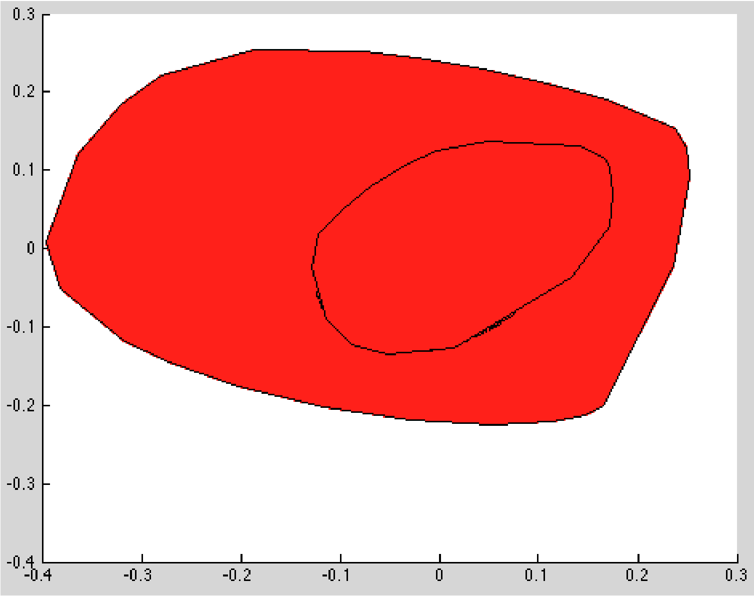
(w/ A. Majumdar, Stanford)

More recent directions…

Iterative DSOS/SDSOS via
- Column generation

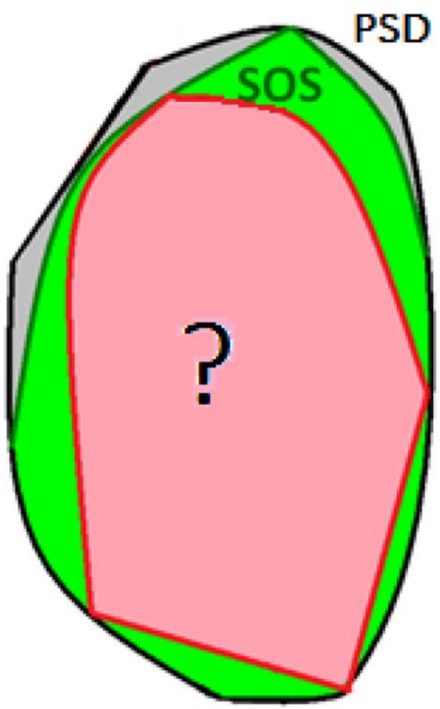
- Cholesky change of basis

(Next talk!)



(w/ S. Dash, IBM,
G. Hall, Princeton)

Main messages



Want to know more?
aaa.princeton.edu

Workshop webpage:
aaa.princeton.edu/largesdps

Backup slides…

r-dsos can in fact outperform sos

The Motzkin polynomial:

$$M(x_1, x_2, x_3) = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 x_3^2 + x_3^6$$

nonnegative but *not* sos!

…but it’s 2-dsos.

(certificate of nonnegativity using LP)

Another ternary sextic:

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

***not* sos, but 1-dsos (hence nonnegative)**

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							

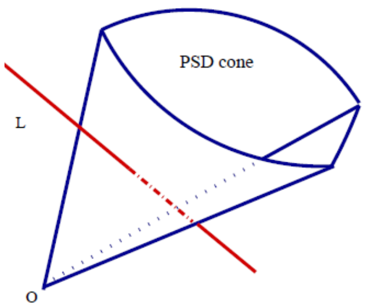
SOS→SDP

Q. Is it any easier to decide sos? **[Lasserre], [Nesterov], [Parrilo]**

- Yes! Can be reduced to a **semidefinite program (SDP)**

Thm: A polynomial ***p(x)*** of degree ***2d*** is sos if and only if there exists a matrix Q such that

$$Q \succcurlyeq 0,$$
$$p(x) = z(x)^T Q z(x),$$



where

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

Application 1: polynomial optimization

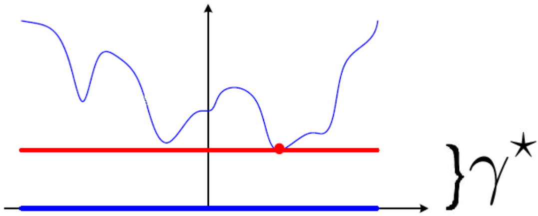
$$\boxed{\begin{aligned} &\min_x p(x) \\ &f_i(x) \leq 0 \\ &h_i(x) = 0 \end{aligned}}$$

Equivalent formulation:

$$\max_\gamma \gamma$$

$$p(x) - \gamma \geq 0$$

$$\forall x \in \{f_i(x) \leq 0, \; h_i(x) = 0\}$$



- **Many applications:**
  - Combinatorial optimization (including all problems in NP)
  - Computation of equilibria in games
  - Machine learning (shape constrained regression, topic modeling, etc.)
  - The optimal power flow (OPF) problem
  - Sensor network localization
  - Optimal configurations for formation flying