Sum of Squares Optimization and Its Applications

Amir Ali Ahmadi
Princeton University
Dept. of Operations Research and Financial Engineering (ORFE)

ORF 523

Optimization over nonnegative polynomials

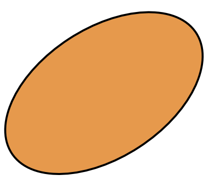
Definition by example: How to pick c_1, c_2, c_3 so to make

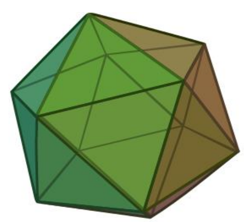
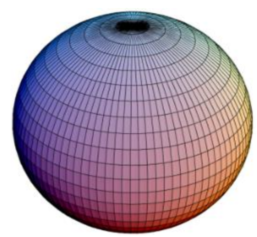
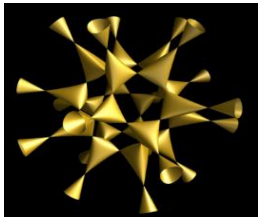
$$p(x_1,x_2) = c_1 x_1^4 - 6x_1^3 x_2 - 4x_1^3 + c_2 x_1^2 x_2^2 + 10x_1^2 + 12x_1 x_2^2 + c_3 x_2^4$$

nonnegative over a given basic semialgebraic set?

Basic semialgebraic set: $\{x \in \mathbb{R}^n |\ g_i(x) \geq 0, h_j(x) = 0\}$

Ex: $x_1^3 - 2x_1 x_2^4 \geq 0$
$x_1^4 + 3x_1 x_2 - x_2^6 \geq 0$



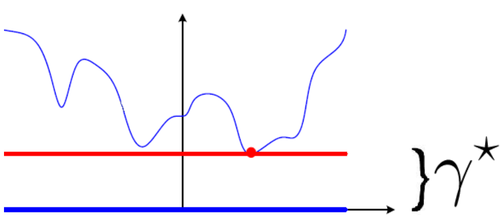
-This problem is fundamental to many areas of applied/computational mathematics.
-It is the problem that "SOS optimization" is designed to solve.

Why would you want to do this?!

- Let's start with five application domains…

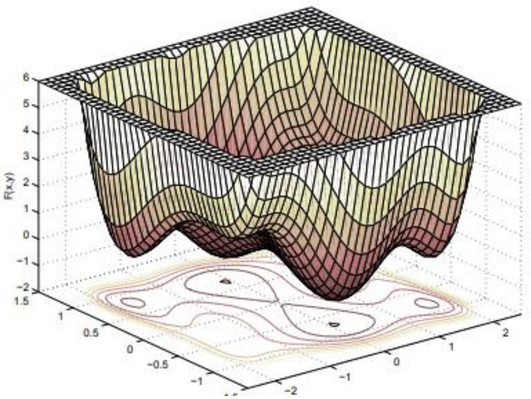
1. Polynomial optimization

$$\left[\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & p(x) \\ \text{s.t.} & g_i(x) \geq 0 \\ & h_j(x) = 0 \end{array}\right] \longleftrightarrow \left[\begin{array}{ll} \max\limits_{\gamma \in \mathbb{R}} & \gamma \\ \text{s.t.} & p(x) - \gamma \geq 0 \quad \forall x \in K \end{array}\right]$$



$p, g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ polynomials

Feasible set $K := \{x \in \mathbb{R}^n \mid g_i(x) \geq 0, h_j(x) = 0\}$



- **Many applications:** the optimal power flow problem, low-rank matrix factorization, dictionary learning, training of deep nets with polynomial activation function, sparse regression with nonconvex regularizes, etc.
- Intractable in general (includes your favorite NP-complete problem)

2. Optimization under input uncertainty

How to make optimal decisions when input to optimization problem is uncertain/noisy?

Example: The Markowitz portfolio optimization problem

$$
\begin{aligned}
\max_{x\in\mathbb{R}^n,\ \gamma\in\mathbb{R}} \quad & \gamma \\
\text{s.t.} \quad & \mu^T x \geq \gamma \quad \text{(return)} \\
& x^T \Sigma x \leq \delta \quad \text{(risk)} \\
& x \geq 0, \ \sum_{i=1}^{n} x_i = 1 \\
& x \in \Omega
\end{aligned}
$$

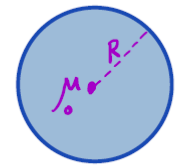
$\mu \in \mathbb{R}^n$: mean vector of the returns

$\Sigma \in \mathbb{S}^{n\times n}$: covariance matrix of the returns

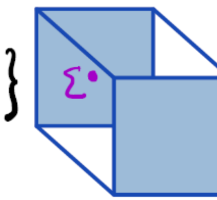
In practice estimated from past data/ML model. Optimal portfolio sensitive to this input.

Accounting for uncertainty:

$$U_\mu = \{ \mu_0 + u \in \mathbb{R}^n \mid \|u\| \leq R \}$$



$$U_\Sigma = \{ \Sigma \in S^{n\times n} \mid \Sigma \succeq 0, \ \Sigma^l_{ij} \leq \Sigma_{ij} \leq \Sigma^u_{ij} \}$$

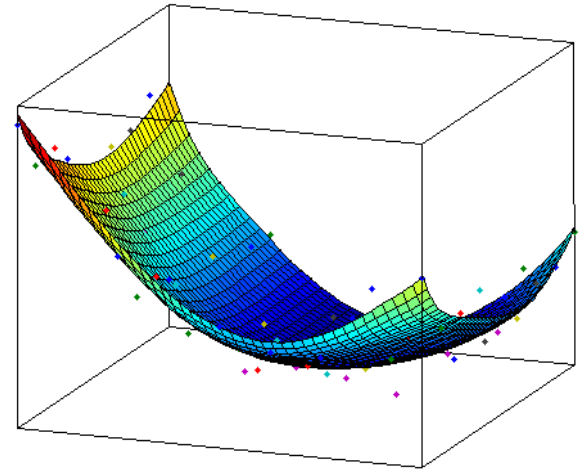


$$
\begin{aligned}
\max_{x\in\mathbb{R}^n,\ \gamma\in\mathbb{R}} \quad & \gamma \\
\text{s.t.} \quad & \mu^T x \geq \gamma \quad \forall \mu \in U_\mu \\
& x^T \Sigma x \leq \delta \quad \forall \Sigma \in U_\Sigma \\
& x \geq 0, \ \sum_{i=1}^{n} x_i = 1 \\
& x \in \Omega
\end{aligned}
$$

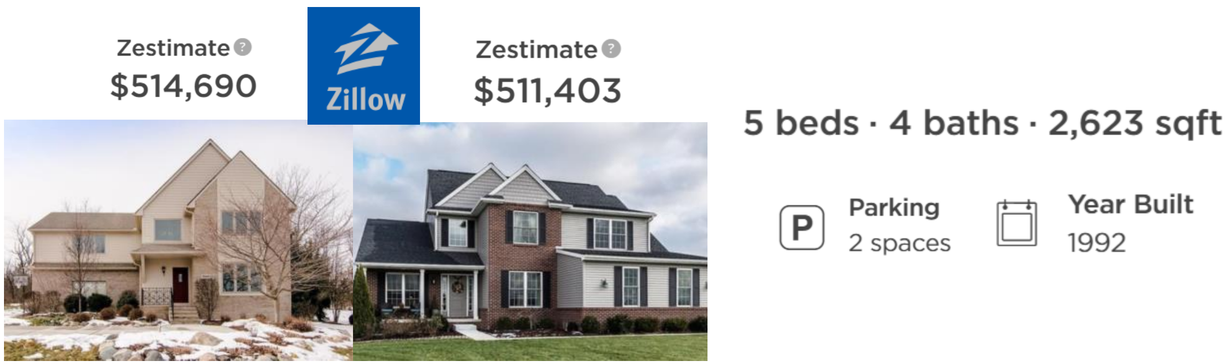
3. Statistics and machine learning

Shape-constrained regression; e.g., *monotone and/or convex regression*

Shape constraints act as regularizer, improve test performance, make model more interpretable and trustworthy



Example 1: Shape constraints natural in most applications



Monotonicity of a polynomial $p(x_1, \dots, x_n)$ with respect to feature j: $\dfrac{\partial p(x)}{\partial x_j} \geq 0, \forall x \in B$

Example 2: "ML for fast real-time convex optimization"

$$
\begin{aligned}
g(b) := \min_{x \in \mathbb{R}^n} \; & f_0(x) \\
\text{s.t.} \; & f_i(x) \leq b_i \; i = 1, \dots, m \\
& x \in \Omega
\end{aligned}
$$

$f_0, \dots, f_m$ convex functions, Ω a convex set.

Goal: learn $g(b)$ offline from training set; evaluate it online very fast

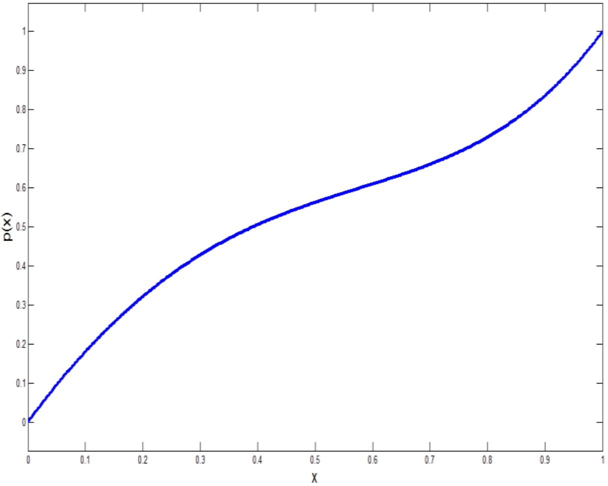
$g: \mathbb{R}^m \to \mathbb{R}$ is
- convex
- nonincreasing w.r.t. all arguments

$y^T \nabla^2 g(b) y \geq 0, \forall b, \forall y$ $\quad \dfrac{\partial g(b)}{\partial b_j} \leq 0, \forall b, \forall j$

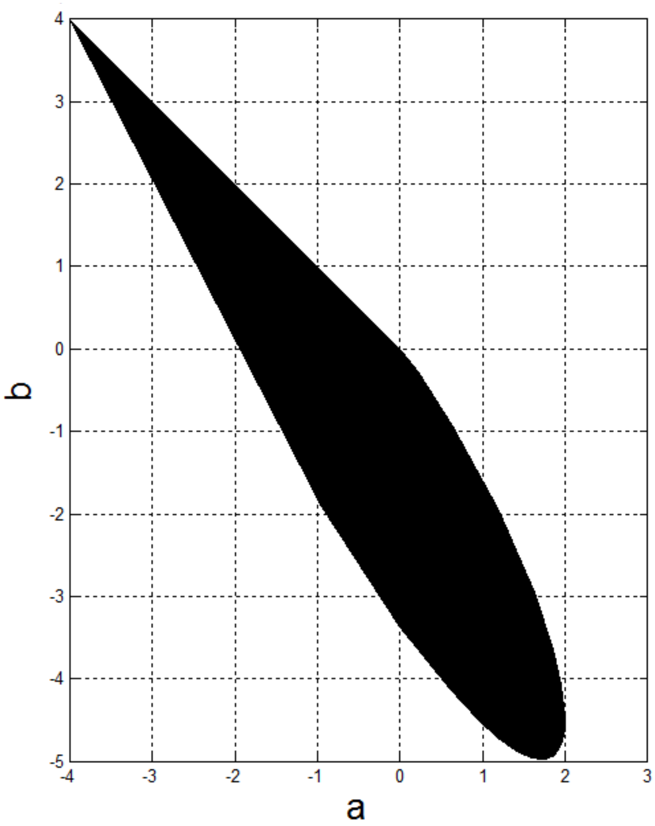
Imposing monotonicity

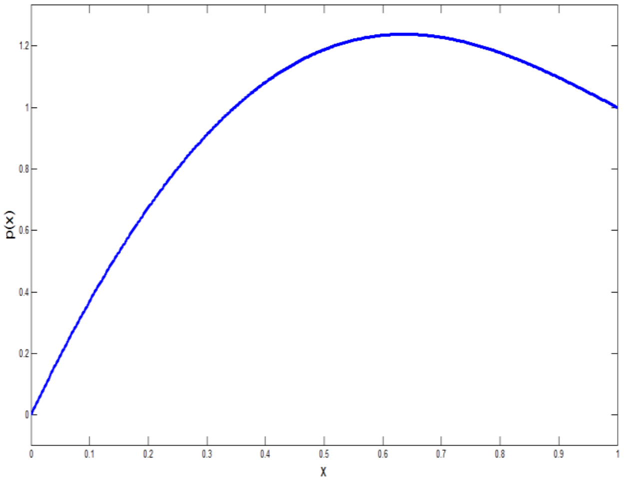
- For what values of a, b is the following polynomial monotone over $[0,1]$?

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



$a = 0, b = -2$

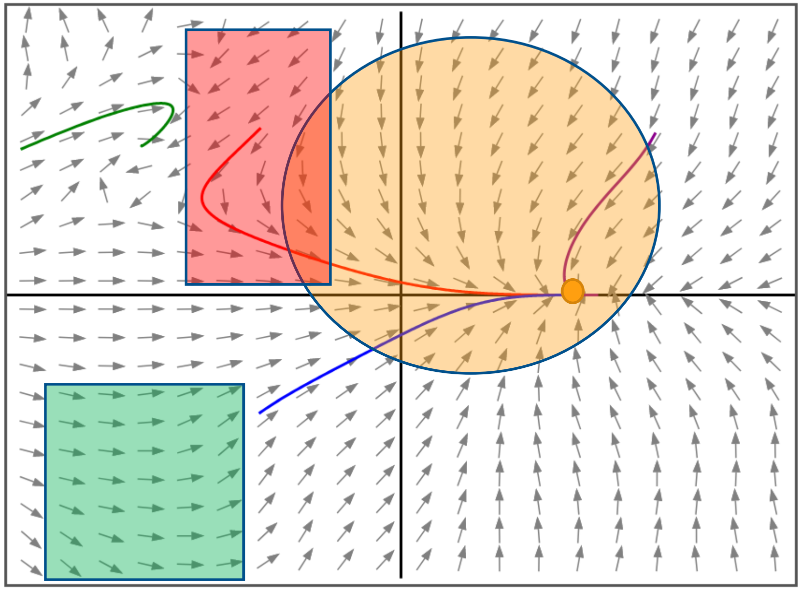




$a = -1, b = -3$

4. Certifying properties of dynamical systems

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



Questions about properties of dynamical systems (e.g., stability, safety)

Lyapunov theory →

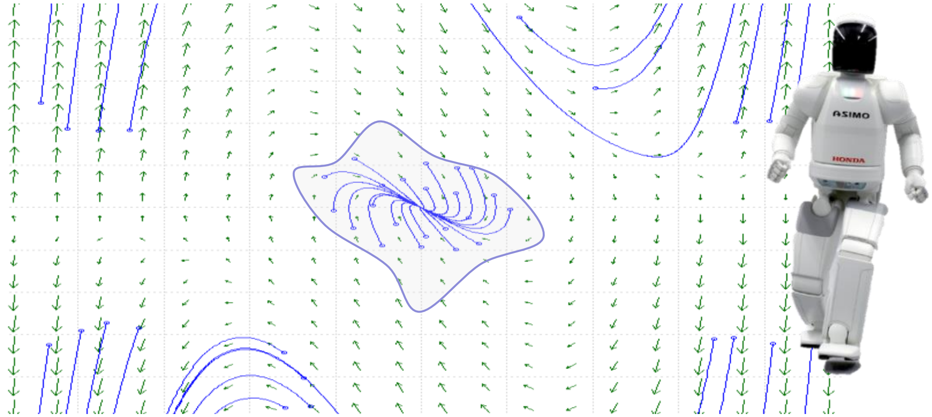
Search for functions satisfying nonnegativity constraints

Semialgebraic parametrization →

Polynomial inequalities

Example: certifying stability

$$\dot{x} = f(x) \quad f : \mathbb{R}^n \to \mathbb{R}^n$$

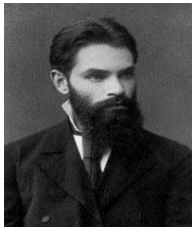


Ex. $\dot{x}_1 = -x_2 + \frac{3}{2}x_1^2 - \frac{1}{2}x_1^3 x_2$

$\dot{x}_2 = 3x_1 - x_1 x_2$

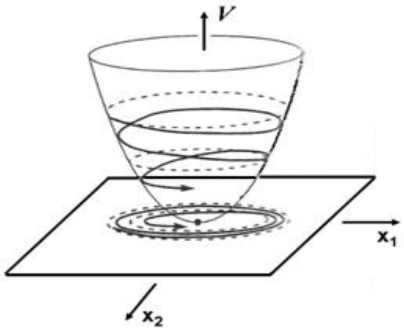
Locally asymptotic stability (LAS) of equilibrium points

Lyapunov's theorem (and its converse):



The origin is LAS if and only if there exists a C^1 function $V: \mathbb{R}^n \to \mathbb{R}$ that vanishes at the origin and a scalar $\beta > 0$ such that

$$V(x) > 0$$
$$V(x) \leq \beta \Rightarrow \dot{V}(x) = \nabla V(x)^T f(x) < 0$$

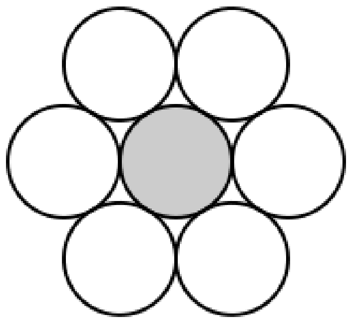


(If $\dot{V}(x) < 0$ everywhere, then globally stable.)

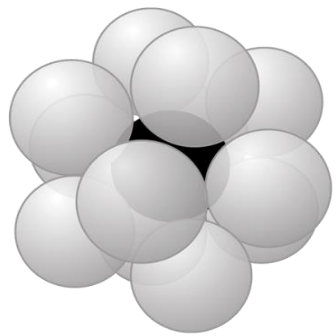
5. Automated theorem proving in geometry

- **Kissing number in dimension n:** largest number of n-dimensional non-overlapping spheres that can simultaneously touch (or "kiss") a common unit sphere.

$k_2 = 6$



$k_3 = ?$



$k_3 \geq 12$

Newton

$k_3 = 12$

Gregory

$k_3 = 13$

Discussion/bet in 1694

Newton proved to be correct in 1953!

13 spheres impossible iff the following system is ***infeasible***:

$$x_i^2 + y_i^2 + z_i^2 = 4, \;\; i = 1, \ldots, 13$$
$$(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 \geq 4,$$
$$i, j \in \{1, \ldots, 13\}^2$$

$$\boxed{\left\{\begin{array}{l} g_1(x) \geq 0 \\ \vdots \\ g_{100}(x) \geq 0 \end{array}\right\} \text{ infeasible}} \Leftrightarrow \boxed{\begin{bmatrix} g_1(x) \geq 0 \\ \vdots \\ g_{99}(x) \geq 0 \end{bmatrix} \Rightarrow g_{100}(x) < 0}$$

Outline of the rest of the talk…

- **Global nonnegativity**
  - Sum of squares (SOS) and semidefinite programming
  - Two applications
  - Hilbert's 17^{th} problem
- **Nonnegativity over a region**
  - Positivstellensatze of Stengle and Putinar
  - Three applications
- **Recap and further reading**

How would you prove nonnegativity?

Ex. Decide if the following polynomial is nonnegative:

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

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

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

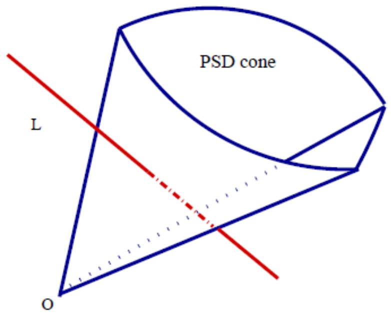
Natural questions:

- **Q1:** Is it any easier to test for a sum of squares (SOS) decomposition?
- **Q2:** Is every nonnegative polynomial SOS?

Sum of squares and semidefinite programming

[Lasserre], [Nesterov], [Parrilo]

Q1: Is it any easier to decide SOS?



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

- Can also efficiently **search and optimize** over SOS polynomials

- As we will see, this latter property is very important in applications…

Semidefinite programming (SDP)

- A broad generalization of linear programs $LP \subseteq (\text{Convex})\ QP \subseteq SOCP \subseteq SDP$
- Can be solved to arbitrary accuracy in polynomial time (e.g., using interior point algorithms) **[Nesterov, Nemirovski], [Alizadeh]**

$$\boxed{\begin{aligned} \min_{X \in S^{n\times n}} \quad & Tr(CX) \\ \text{s.t.} \quad & Tr(A_i X) = b_i \quad i=1,\dots,m \\ & X \succeq 0 \ \leftarrow \text{"psd"} \end{aligned}}$$

Data to SDP: $C, A_1, \dots, A_m \in S^{n\times n}, b_1, \dots, b_m \in \mathbb{R}$

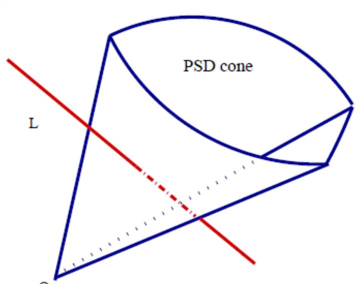
Notes: $Tr(CX) = \sum_{i,j} C_{ij} X_{ij}$

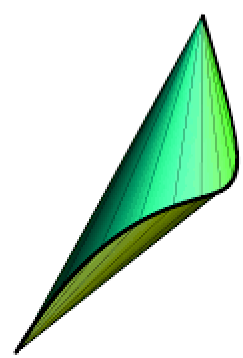
$X \succeq 0: \quad y^T X y \geq 0 \quad \forall y \in \mathbb{R}^n$

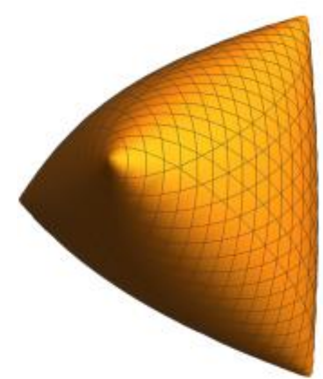
$\Updownarrow$

Eigenvalues of X are ≥ 0.

Feasible set called a "spectrahedron":







SOS→SDP

Thm:

A polynomial p of degree $2d$ is SOS if and only if $\exists Q \succcurlyeq 0$ such that

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

where $z = [1, x_1, \dots, x_n, x_1x_2, \dots, x_n^d]^T$ is the vector of monomials of degree up to d.

(It follows that checking membership or optimizing a linear function over the set of SOS polynomials is an SDP)

Proof: $(\Rightarrow)$ Suppose $\exists Q \succcurlyeq 0$ s.t. $p(x) = z^T(x) Q z(x)$ $\forall x$.

$$Q \succcurlyeq 0 \Rightarrow Q = \underbrace{V^T V}_{V:\ r \times \binom{n+d}{d}} \Rightarrow p(x) = z^T(x) V^T V z(x) = \| V z(x) \|^2 = \sum_{i=1}^{r} (v_i^T z(x))^2.$$

$(\Leftarrow)$ Suppose $p(x)$ is SOS.

$\exists v_1, \dots, v_r \in \mathbb{R}^{\binom{n+d}{d}}$ s.t. $p(x) = \sum_{i=1}^{r} (v_i^T z(x))^2 = \sum_{i=1}^{r} (z^T(x) v_i)(v_i^T z(x)) = z^T(x) \overbrace{\left(\sum_{i=1}^{r} v_i v_i^T \right)}^{:= Q} z(x).$

Example

$$\boxed{p(x) = 10x^4 - 2x^3 - 7x^2 + 4x + 4}$$

Is p SOS ?

SDP feasibility problem

$$p(x) = \underbrace{\begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}^T}_{z^T(x)} \underbrace{\begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}}_{z(x)}$$

$\forall x$

Find $Q \succeq 0$ s.t.

$q_{33} = 10$, $2q_{23} = -2$

$q_{22} + 2q_{13} = -7$

$2q_{12} = 4$, $q_{11} = 4$

SDP solver output:

$$Q = \begin{bmatrix} 4 & 2 & -6 \\ 2 & 5 & -1 \\ -6 & -1 & 10 \end{bmatrix} \overset{\text{e.g. Cholesky}}{=} \underbrace{\begin{bmatrix} 0 & 2 \\ 2 & 1 \\ 1 & -3 \end{bmatrix}}_{V^T} \underbrace{\begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & -3 \end{bmatrix}}_{V}$$

$$\Rightarrow p(x) = z^T(x) V^T V z(x) = \| V z(x) \|^2 = \left\| \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \right\|^2$$

$$\boxed{p(x) = (2x + x^2)^2 + (2 + x - 3x^2)^2}$$

Let's revisit two of our applications!

Optimization over nonnegative polynomials

↑

Sum of squares (SOS) programming

↑

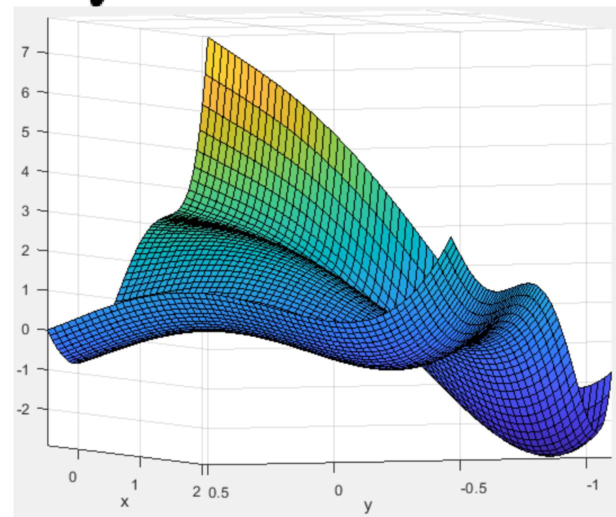
Semidefinite programming (SDP)

1) Nonconvex unconstrained minimization

Find: $$p^* := \inf_{(x,y)\in\mathbb{R}^2} \; 4x^2 - \frac{21}{10}x^4 + \frac{1}{3}x^6 + xy - 4y^2 + 4y^4 + x^2y \;\longleftarrow\; p(x,y)$$

$$\boxed{\begin{array}{lll} p_{sos} := & \sup\limits_{\gamma\in\mathbb{R}} & \gamma \\ & \text{s.t.} & p(x,y) - \gamma \quad \text{sos} \end{array}} \rightarrow \text{SDP!}$$

$$p_{sos} \leq p^*$$




```
p=4*x^2-2.1*x^4+(1/3)*x^6+1*x*y-4*y^2+4*y^4+x^2*y;
solvesos(sos(p-gam),-gam,[],[gam])
p_sos=double(gam)
```

solvertime: 0.6 (s)

p_sos =

-2.921560950963582

```
[inf,z,Q]=solvesos(p-p_sos);
sdisplay(z{1})
[v,d]=eig(double(Q{1}));
zxstar=v(:,1)/v(1,1);
xstar=[zxstar(3);zxstar(2)]
p_at_xstar=replace(p,[x,y],[xstar(1),xstar(2)])
```

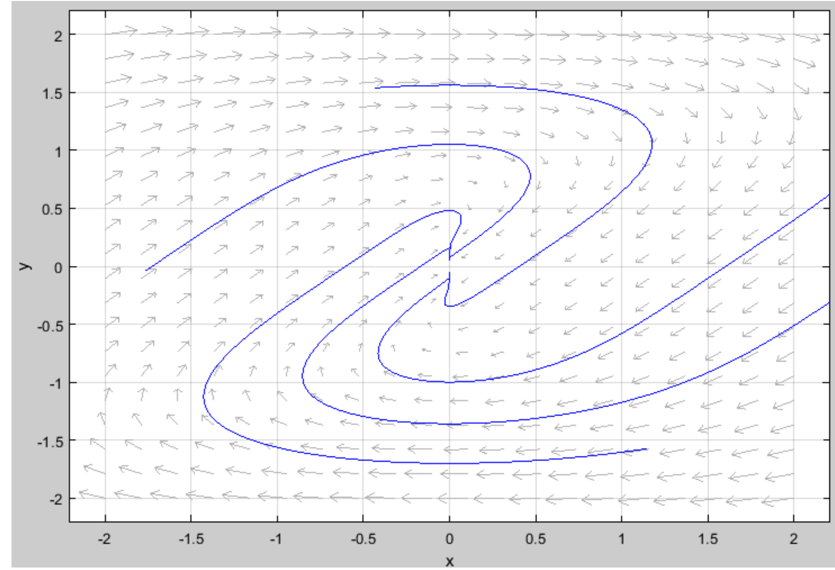
```
xstar =

   1.83299614475561
  -0.922931478421273

p_at_xstar =

  -2.921559422066406
```

2) Automated proof of global asymptotic stability



MK [redacted].edu> on behalf of [redacted]

Tue 2/9/2021 1:13 PM

To: Amir Ali Ahmadi

Hi Amir Ali,

I hope life and career are going well.

I have a question that I assume might take little more than 5-10 of your time but please feel free to let me know if it would actually take more.

> Today in class we got into an interesting discussion with students about what a strict Lyapunov function would be for the system
>
> dx/dt = -x +y^3
> dy/dt = -x
>
> A non-strict L.f.. is easy, V = x^2/2 + y^4/4, with dV/dt = - x^2. One could then deduce g.a.s. by a Barbashin-Krasovskii/Lasalle argument, but that's not satisfactory.
>
> I started constructing a strict one in real time and it quickly got out of hand, necessitating higher and higher powers and many cross terms. I inevitably thought of you and your (and Pablo's) SOS program that would spit out a good strict V within seconds.

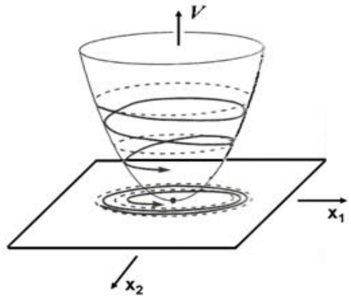
If you can plug in this system and let me know what comes out, I'd appreciate it, and my 40-50 students in class would learn a few things (complexity of Lyapunov functions, automated options for finding them, etc.).

Best regards,

[redacted]

Automated proof of global asymptotic stability

$$\begin{bmatrix} \dot{x} = -x + y^3 \\ \dot{y} = -x \end{bmatrix}$$



Find $V(x,y)$ of degree 4 s.t.

$V(x,y)$ SOS

$-\dot{V}(x,y) = -\left\langle \nabla V(x,y), \begin{pmatrix} -x+y^3 \\ -x \end{pmatrix} \right\rangle$ SOS

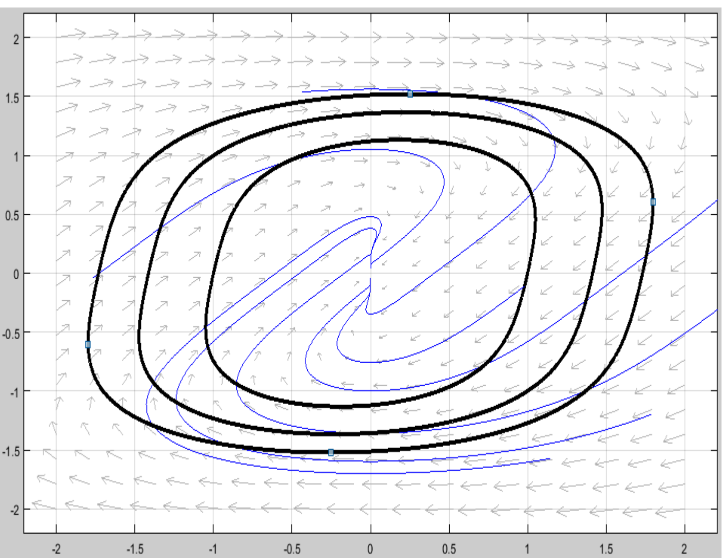
$\rightarrow$ SDP!

```
sdpvar x y
xdot=-x+y^3;
ydot=-x;

[V,c,m]=polynomial([x;y],4,2);
Vdot=jacobian(V,[x,y])*[xdot;ydot];
FF=[sos(V),sos(-Vdot)]
solvesos(FF,[],[],[c])
```

```
>> sdisplay(clean(double(c)'*m,1e-3))
1.00000084865*x^2-0.333330248293*x*y+0.166665124147*y^2+0.500118639025*y^4
```

$$V(x,y) = x^2 - \frac{1}{3}xy + \frac{1}{6}y^2 + \frac{1}{2}y^4$$



$$= \left(x - \frac{1}{6}y\right)^2 + \frac{5}{36}y^2 + \frac{1}{2}y^4 \quad \text{(hence positive definite)}$$

$$\geq \frac{1}{36}(x^2+y^2) + \frac{1}{2}y^4 \quad \text{(hence radially unbounded)}$$

$$\dot{V}(x,y) = -\frac{5}{3}x^2 - \frac{1}{3}y^4 \quad \text{(hence negative definite)}$$

Hilbert's 1888 Paper

Q2: SOS $\overset{?}{\Leftarrow}$ Nonnegativity

n,d	2	4	≥6
1	yes	yes	yes
2	yes	yes	no
3	yes	no	no
≥4	yes	no	no

From Logicomix

Motzkin (1967):

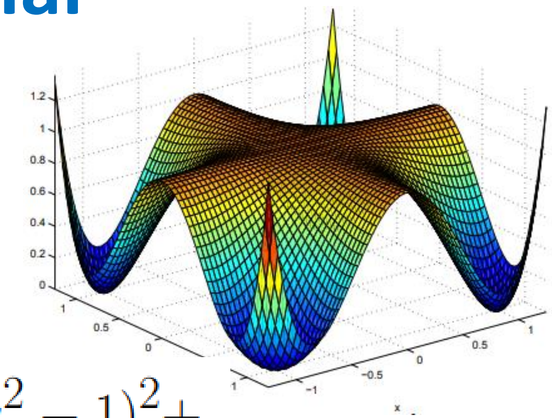
$$M(x_1,x_2) = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$$

Robinson (1973):

$$R(x_1,x_2,x_3) = x_1^2(x_1-1)^2 + x_2^2(x_2-1)^2 + x_3^2(x_3-1)^2 + 2x_1x_2x_3(x_1+x_2+x_3-2)$$

The Motzkin polynomial

$$M(x,y) = x^2y^4 + x^4y^2 + 1 - 3x^2y^2$$

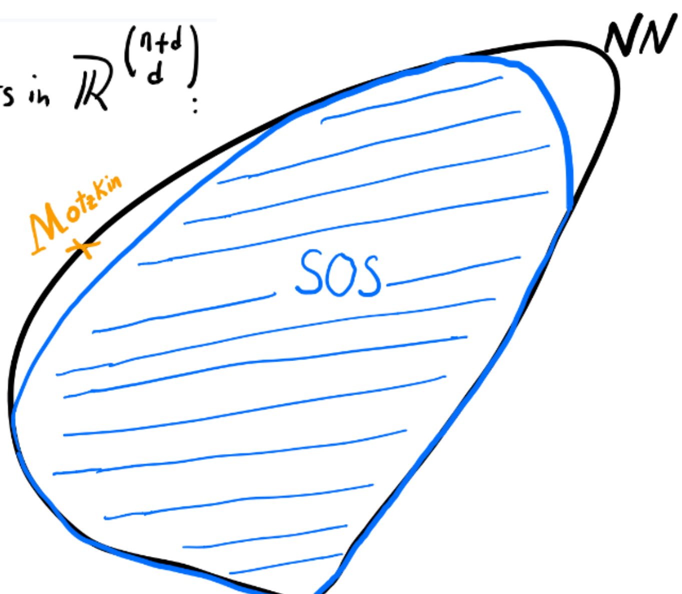


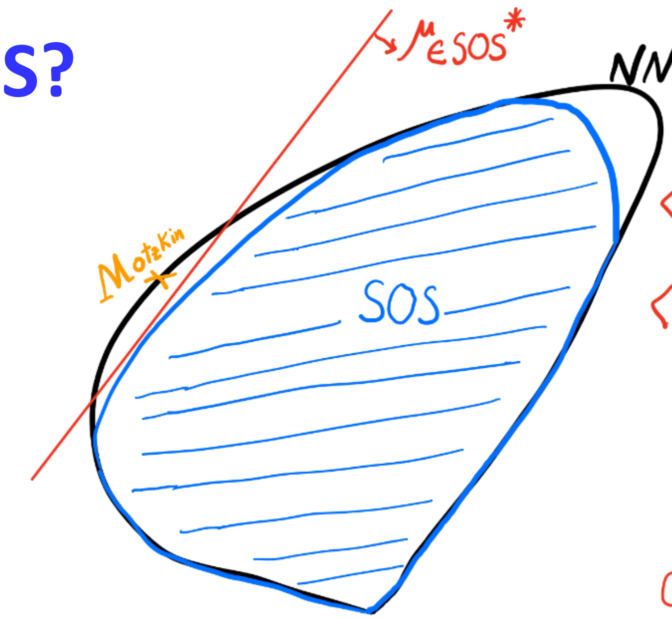
How to prove it is nonnegative?

$$(x^2+y^2+1)\,M(x,y) = (x^2y-y)^2 + (xy^2-x)^2 + (x^2y^2-1)^2 + \frac{1}{4}(xy^3-x^3y)^2 + \frac{3}{4}(xy^3+x^3y-2xy)^2$$

How to prove it is not SOS?

Two sets in $\mathbb{R}^{\binom{n+d}{d}}$:





$\langle \mu, \mathrm{coeff}(M) \rangle < 0$

$\langle \mu, \mathrm{coeff}(q) \rangle \geq 0 \;\; \forall q \in SOS$

$\Updownarrow$

$\left[z(x) z(x)^T \right] \Big|_{\mu} \succeq 0$

Can find μ with SDP.

$NN^* \subseteq SOS^*$

Hilbert's 17th Problem (1900)

Q. $\boldsymbol{p}$ **nonnegative** $\overset{?}{\Rightarrow}$ $$p = \sum_i \left(\frac{g_i}{q_i}\right)^2$$

- Artin (1927): **Yes!**

- Implications:
  - $p \geq 0 \Rightarrow \exists \boldsymbol{h}$ **sos** such that $\boldsymbol{p}.\boldsymbol{h}$ **sos**
  - **Reznick:** (under mild conditions) can take $\boldsymbol{h} = \left(\sum_i \boldsymbol{x}_i^2\right)^{\boldsymbol{r}}$
  - Certificates of nonnegativity can *always* be given with sos (i.e., with semidefinite programming)!
  - We'll see how the Positivstellensatz generalizes this even further…

Outline of the rest of the talk…

- **Global nonnegativity**
  - Sum of squares (SOS) and semidefinite programming
  - Two applications
  - Hilbert's 17th problem
- **Nonnegativity over a region**
  - Positivstellensatze of Stengle and Putinar
  - Three applications
- **Recap and further reading**

Positivstellensatz

$p(x) > 0, \forall x \in \mathbb{R}^n$

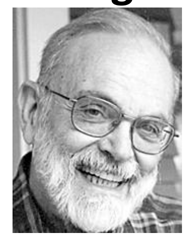
Artin

If $p(x) \geq 0, \forall x \in \mathbb{R}^n$,
then $\exists$ sos q s.t. $p \cdot q$ sos.

1927

20th century

Stengle



1974

1991

1993

$p(x) > 0,$
$\forall x \in S = \{x \mid g_i(x) \geq 0\}$

Schmudgen

Putinar

Positivstellensatz: a complete algebraic proof system

- Let's motivate it with a toy example:

Consider the task of proving the statement:

$$\forall a, b, c, x, \;\; ax^2 + bx + c = 0 \Rightarrow b^2 - 4ac \geq 0$$

Short algebraic proof (certificate):

$$b^2 - 4ac = (2ax + b)^2 - 4a(ax^2 + bx + c)$$

- The Positivstellensatz vastly generalizes what happened here:
  - Algebraic certificates of infeasibility of any system of polynomial inequalities (or algebraic implications among them)
  - **Automated** proof system (via semidefinite programming)

Positivstellensatz: a generalization of Farkas lemma

Farkas lemma (1902):

$Ax = b$ and $x \geq 0$ is infeasible

$$\Updownarrow$$

There exists a y such that $y^T A \geq 0$ and $y^T b < 0$.

(The S-lemma is also a theorem of this type for quadratics)

Stengle's Positivstellensatz (1974)

$$S = \{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\} \quad \textit{is empty}$$

if and only if

there exist sum of squares polynomials $s_0(x), s_1(x), \ldots, s_m(x), s_{12}(x), s_{13}(x), \ldots, s_{123\ldots m}(x)$

such that

$$-1 = s_0(x) + \sum_i s_i(x)g_i(x) + \sum_{\{i,j\}} s_{ij}(x)g_i(x)g_j(x) + \ldots + s_{123\ldots m}(x)g_1(x)\ldots g_m(x).$$

- This is algebraic certificate of emptiness of *S*
- Works in full generality (no assumptions on *S)*
- Degree bounds on SOS multipliers based on $n, m, \deg(g_i)$ only
- Artin's solution to Hilbert's 17th problem is a corollary
- Leads to an SDP hierarchy for polynomial optimization (the ``Parrilo hierarchy'')

Putinar's Positivstellensatz (1993)

$$\boldsymbol{p(x) > 0 \text{ on } S = \{x \in \mathbb{R}^n \mid g_i(x) \geq 0, i = 1, \dots, m\}}$$

easy direction $\Uparrow$ $\Downarrow$ under the "Archimedean condition"
(slightly stronger than compactness of S)

$\exists \epsilon > 0$ and SOS polynomials $s_0(x), \dots, s_m(x)$ such that

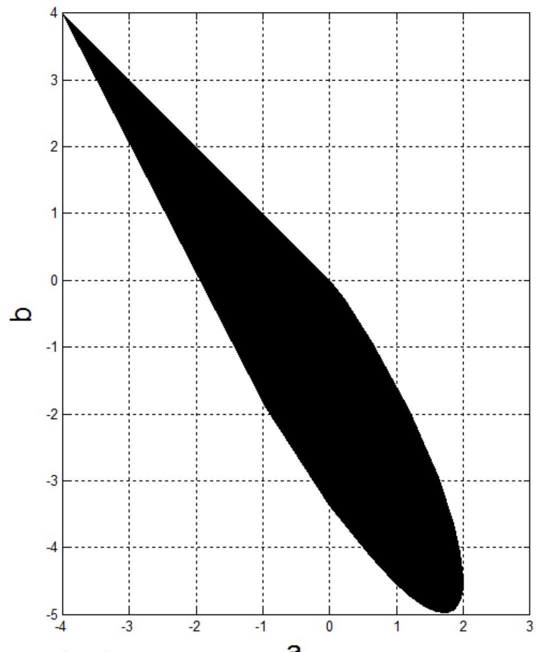
$$\boldsymbol{p(x) - \epsilon = s_0(x) + \textstyle\sum_i s_i(x) g_i(x).}$$

- This is algebraic certificate of positivity
- Leads to an SDP hierarchy for polynomial optimization (the ``Lasserre hierarchy'')
- Degree bounds on SOS multipliers based on the coefficients (though in special cases, better degree bounds possible)

How did I plot this?

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

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



Theorem. A polynomial $p(x)$ of degree $2d$ is monotone on $[0,1]$ if and only if

$$p'(x) = xs_1(x) + (1-x)s_2(x),$$

where $s_1(x)$ and $s_2(x)$ are some SOS polynomials of degree $2d-2$.

Let's end with 3 applications:

- **Finance**
- **Control**
- **Learning dynamical systems**

Optimization over nonnegative polynomials

↑

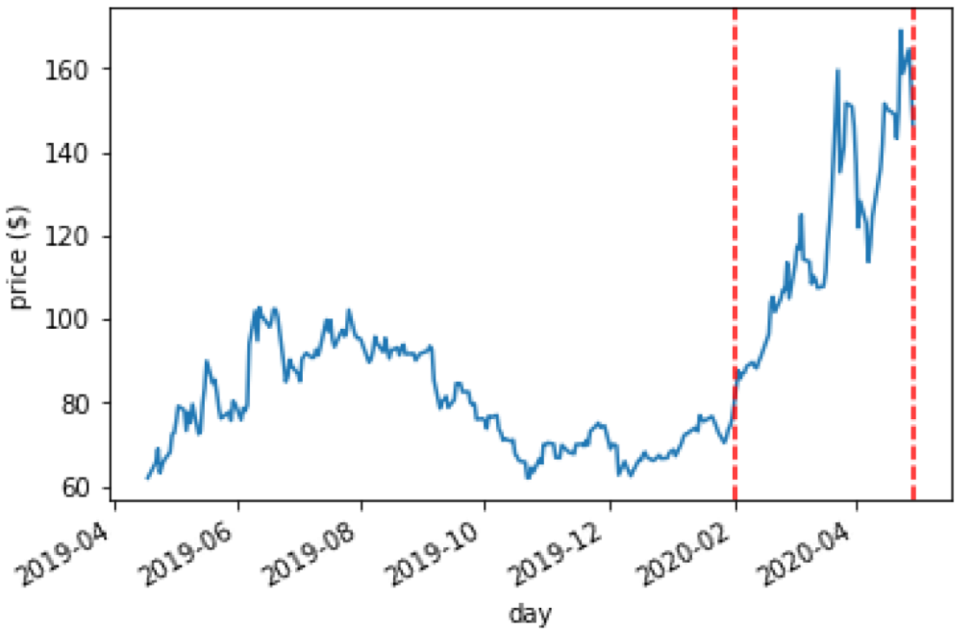
Sum of squares (SOS) programming

↑

Semidefinite programming (SDP)

Distributionally robust optimization

What's the probability that Zoom's stock goes bust?



- Three months starting Feb 1, 2020

$$r_i = \frac{P_i - P_{i-1}}{P_i}, \quad i = 1, \ldots, 61$$

- Empirical moments $m_k = \mathbb{E}[r^k]$:

$$m_1 = 0.0068, m_2 = 0.0034,$$
$$m_3 = 2 \times 10^{-6}, m_4 = 5 \times 10^{-5}$$

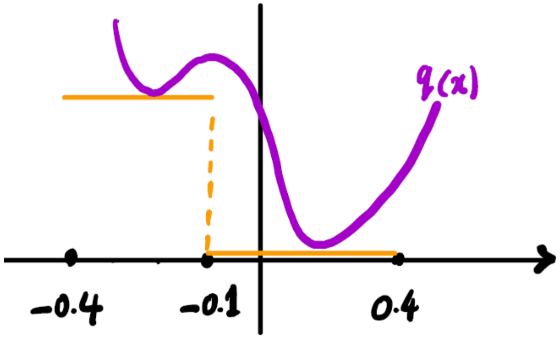
- The distribution of r is supported on [-0.4,0.4] but is otherwise unknown
- What is the probability that Zoom's stock return will be below -0.1 today?
- Want the worst-case probability over all distributions whose first 4 moments are within 10% of those computed from data.

Sum of squares optimization can compute this probability!

$$
\begin{aligned}
\alpha := \inf_{q,r,s,\gamma} \quad & \gamma \\
\text{s.t.} \quad & q(x) = \sum_{k=0}^{4} q_k x^k \ \text{ is a degree-4 (univariate) polynomial,} \\
& r(x), s(x) \ \text{ are quadratic polynomials that are sos,} \\
& q_0 + \sum_{k=1}^{4} q_k m_k' \leq \gamma \ \forall m_k' \in [0.9\, m_k, 1.1\, m_k] \text{ for } k = 1, \ldots, 4, \\
& q(x) - (0.4^2 - x^2)\, s(x) \text{ is sos,} \\
& q(x) - 1 - (0.4 + x)(-0.1 - x) r(x) \text{ is sos.}
\end{aligned}
$$

$$\Longrightarrow q(x) \geq 0 \quad \forall x \in [-0.4, 0.4]$$

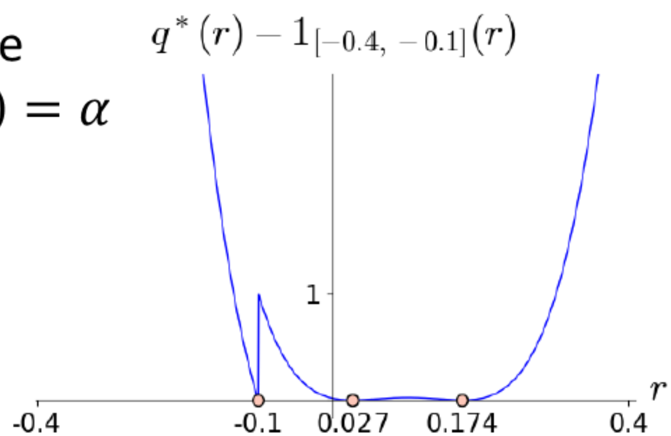
$$\Longrightarrow q(x) \geq 1 \quad \forall x \in [-0.4, -0.1]$$



$$\mathbb{P}(r \in [-0.4, -0.1]) = \mathbb{E}[1_{[-0.4,-0.1]}] \qquad \Rightarrow 1_{[-0.4,-0.1]} \leq q(x) \ \forall x \in [-0.4, 0.4]$$

$$\Rightarrow \mathbb{E}\left[1_{[-0.4,-0.1]}\right] \leq \mathbb{E}[q(x)] = \sum_{k=0}^{4} q_k m_k \leq \gamma$$

In fact, we always have
$\mathbb{P}(r \in [-0.4, -0.1]) = \alpha$



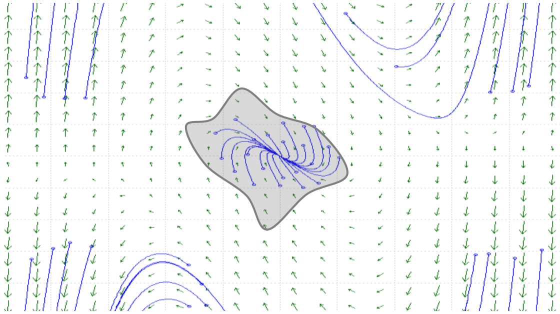
$$\mathbb{P}(r \in [-0.4, -0.1]) \leq \alpha$$

```
Optimizer terminated. Time: 0.17
alpha =

    0.2073
```

SOS proofs of local asymptotic stability (LAS)

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

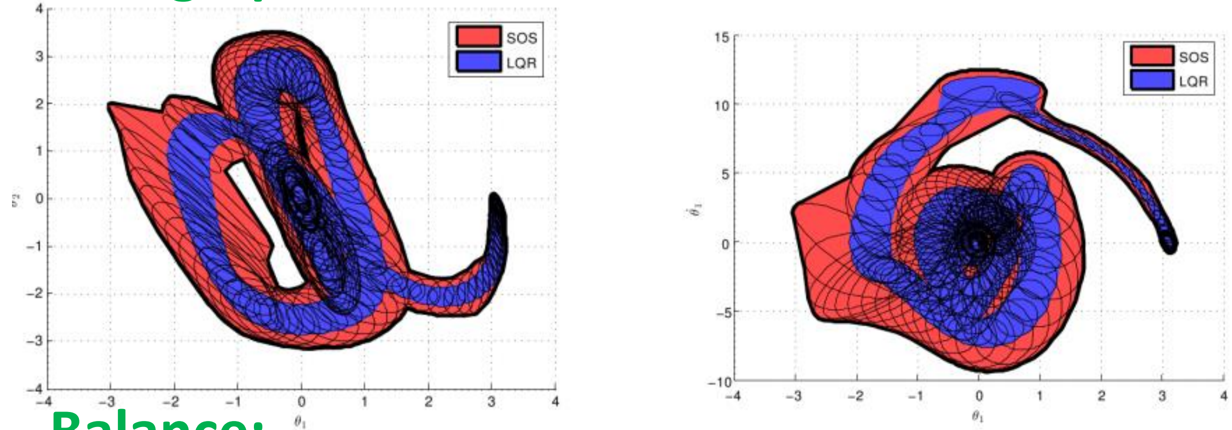


$$V(x) > 0$$
$$V(x) \leq \beta \Rightarrow \dot{V}(x) = \nabla V(x)^T f(x, u) < 0$$

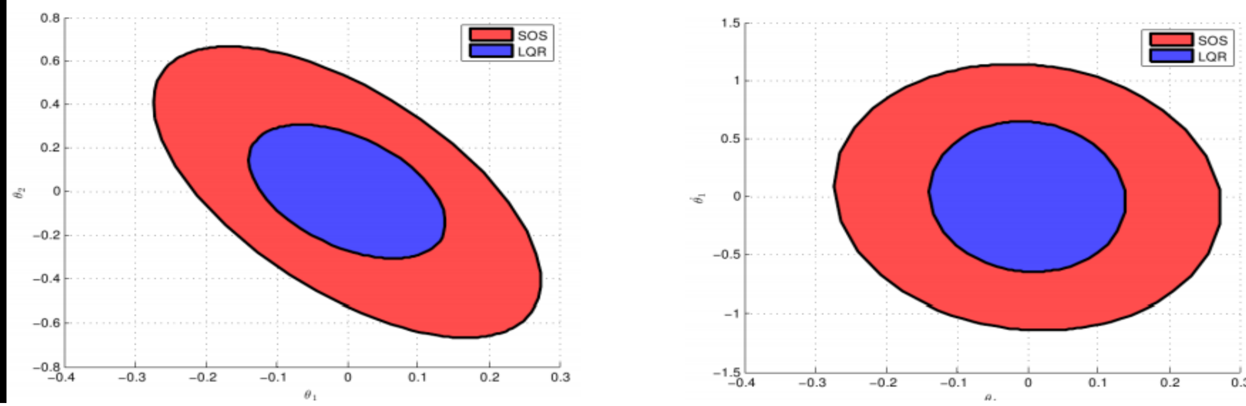
- Deals with nonlinear systems directly.
- Gives easily-verifiable proofs of stability in a fully-automated fashion.

Local stability – SOS on the Acrobot

Swing-up:



Balance:

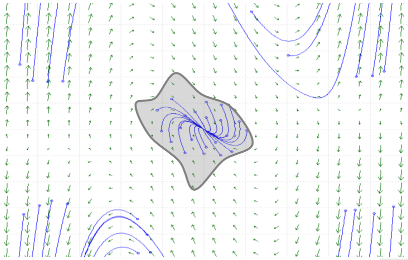


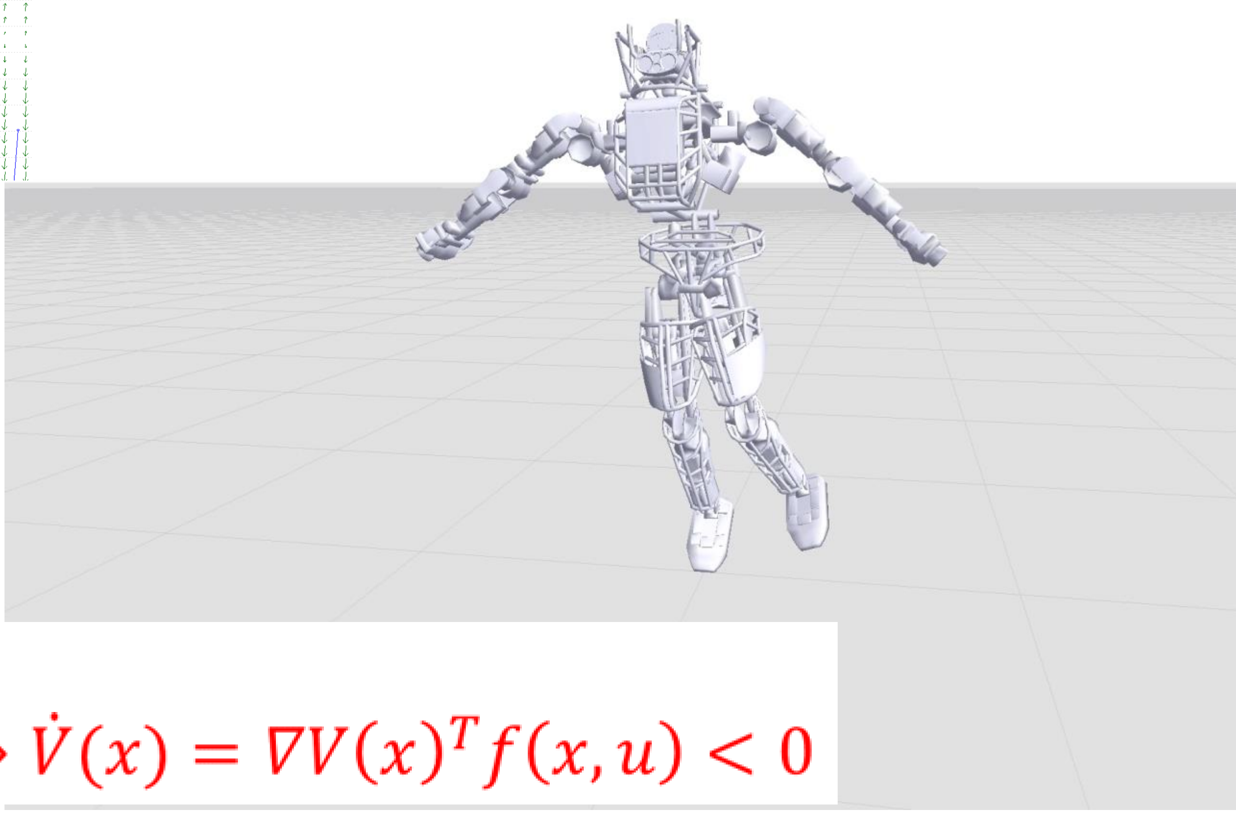
Controller designed by SOS

(w/ Majumdar and Tedrake)

Stabilizing a humanoid robot on one foot

$\dot{x} = f(x,u)$ 30 states 14 control inputs Cubic dynamics





$V(x) > 0$

$V(x) \leq \beta \Rightarrow \dot{V}(x) = \nabla V(x)^T f(x,u) < 0$

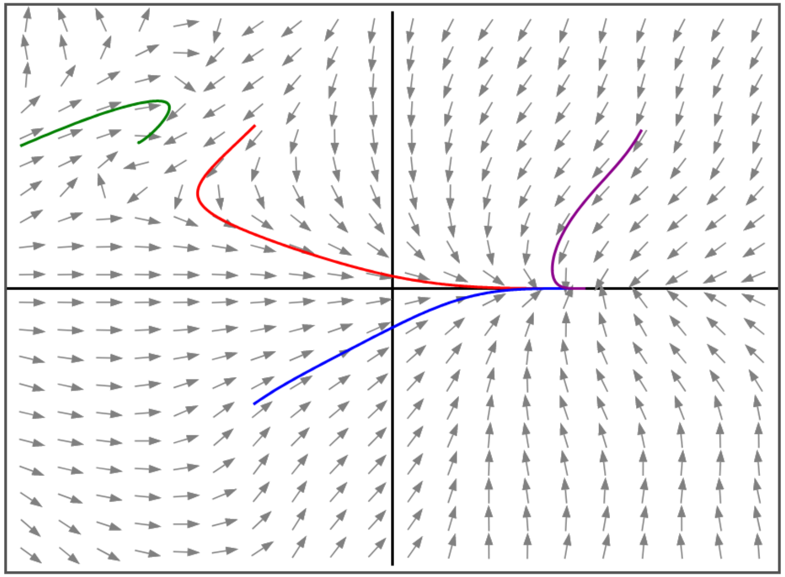
(w/ Majumdar and Tedrake)

Learning dynamical systems with side information

- Goal is to learn a dynamical system

  $$\dot{x} = f(x) \quad (\text{where } f\colon \mathbb{R}^n \to \mathbb{R}^n)$$

  from a *limited* number of *noisy* measurements of its trajectories.



Examples of "side information":

- Equilibrium points (and their stability)
- Invariance of certain sets
- Decrease of certain energy functions
- Sign conditions on derivatives of states
- Having gradient structure
- Monotonicity conditions
- Incremental stability
- (Non)reachability of a set B from a set A, …

- Parametrize a polynomial vector field $p\colon \mathbb{R}^n \to \mathbb{R}^n$.
- Use SOS optimization to impose side information as constraints on p.
- Pick the p that best explains the data.

An epidemiology example

A model from the epidemiology literature for spread of Gonorrhea in a heterosexual population:

$$\dot{x} = f_1(x, y) = -a_1 x + b_1(1 - x)y$$

$$\dot{y} = f_2(x, y) = -a_2 y + b_2(1 - y)x$$

$x(t)$: fraction of infected males at time t

$y(t)$: fraction of infected females at time t

a_1: recovery rate of males

a_2: recovery rate of females

b_1: infection rate of males

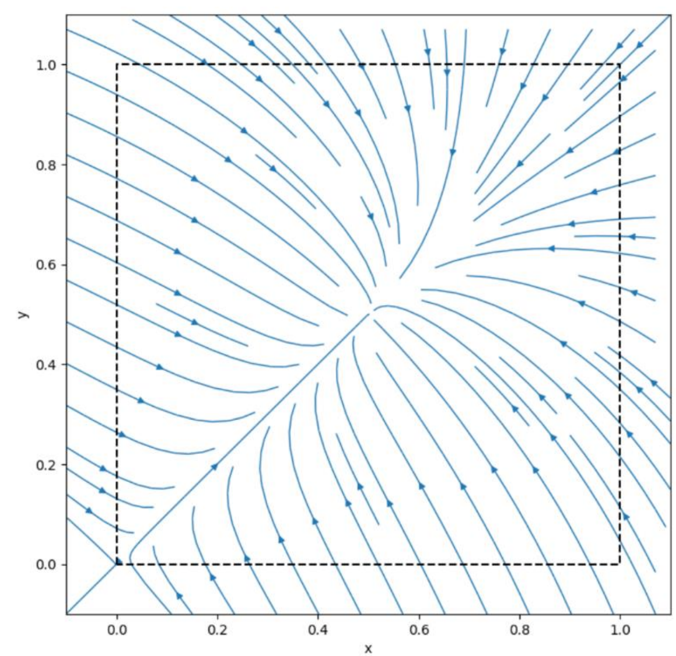
b_2: infection rate of females

For our experiments:

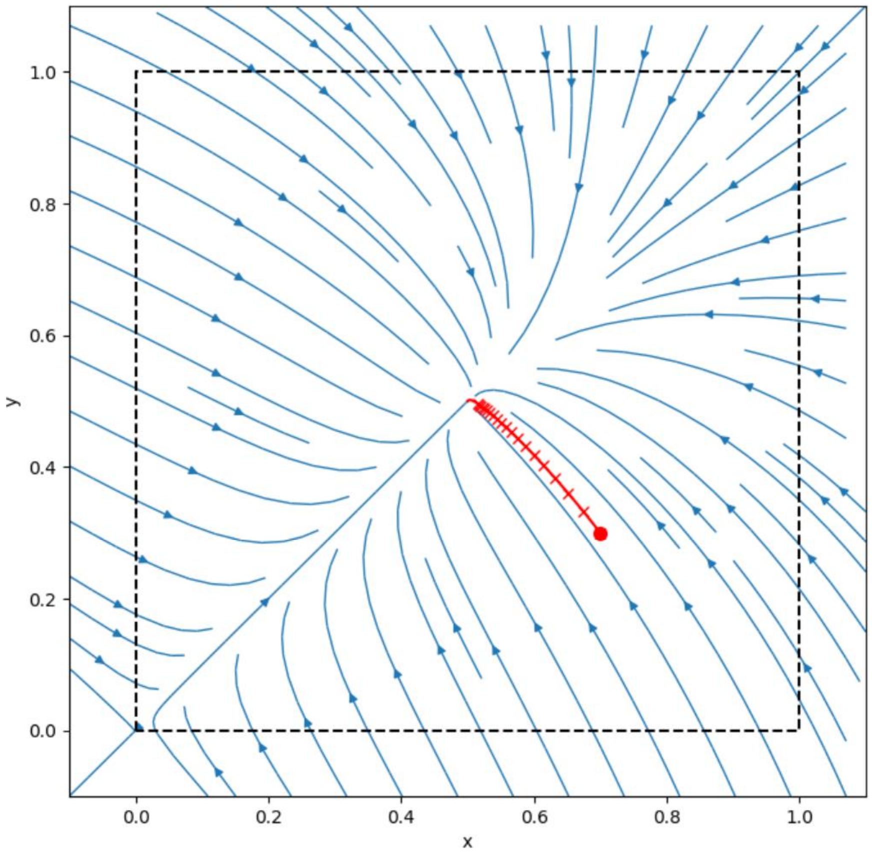
$a_1 = a_2 = .1; b_1 = b_2 = .05.$

This is taken to be "the ground truth".

- The dynamics (both its parameters and its special structure) is unknown to us.
- We only get to observe noisy trajectories of this dynamical system.



The setup



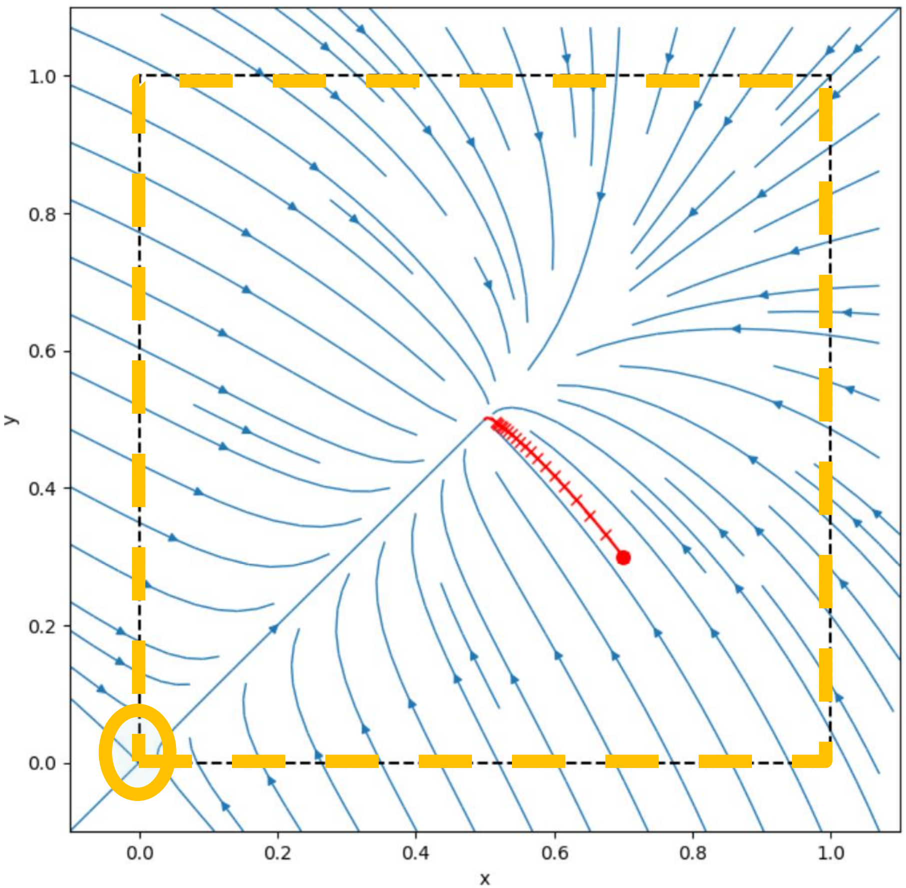
- The true dynamics f is unknown

- **What we observe:**
  Noisy measurements of the vector field on 20 points from a single trajectory starting from [0.7;0.3]

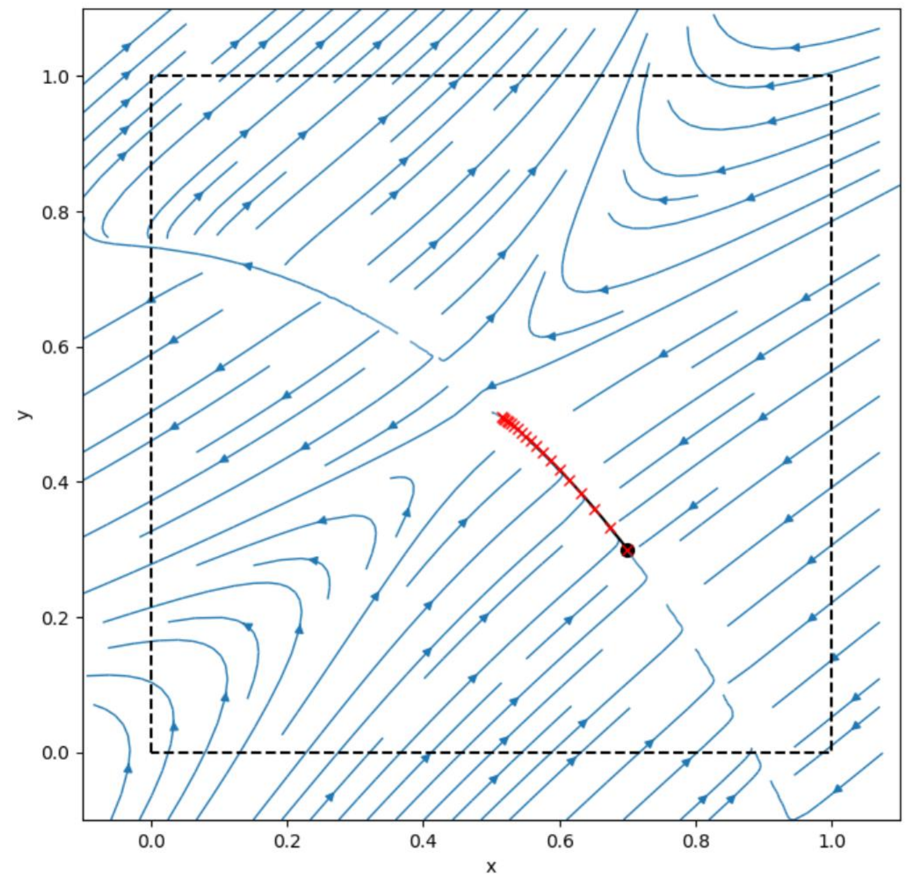
- **Goal:**
  - Learn a polynomial vector field p that best agrees with the observed trajectory
  - Incorporate side information to generalize better to unobserved trajectories

Learning p of degree 3

The true dynamics f (unknown)



Least squares solution



f(0)=0

- Good performance on the observed trajectory. Terrible elsewhere.

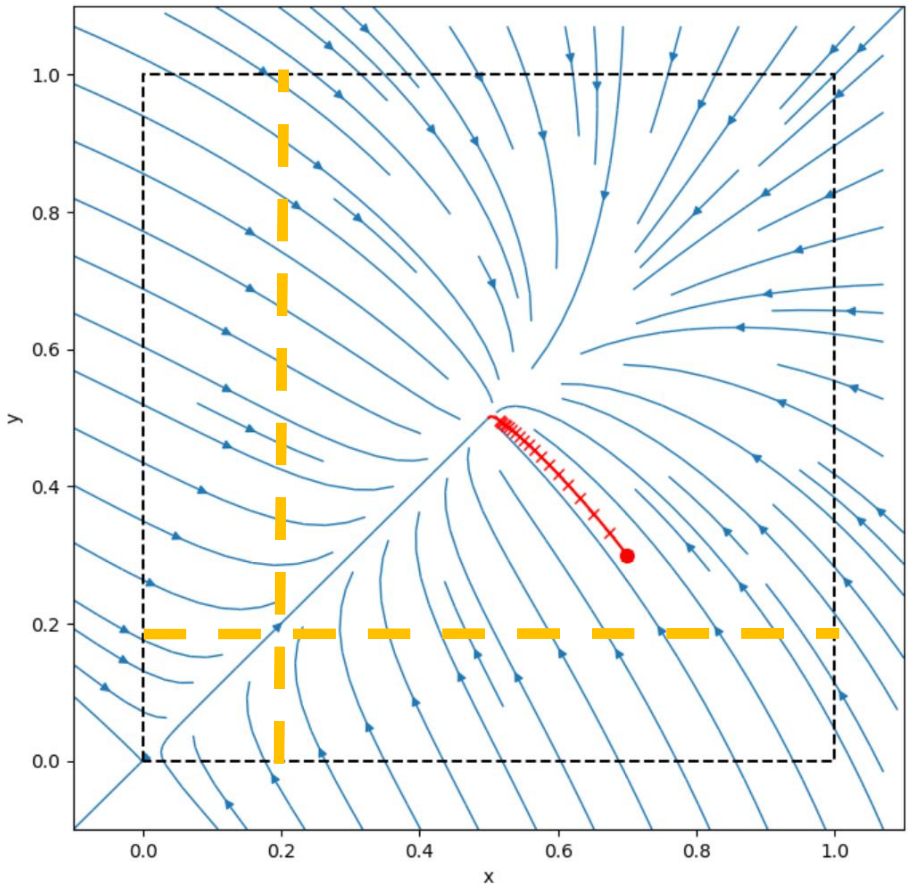
Fraction of infected individuals cannot go negative or more than one!

PRINCETON UNIVERSITY ORFE

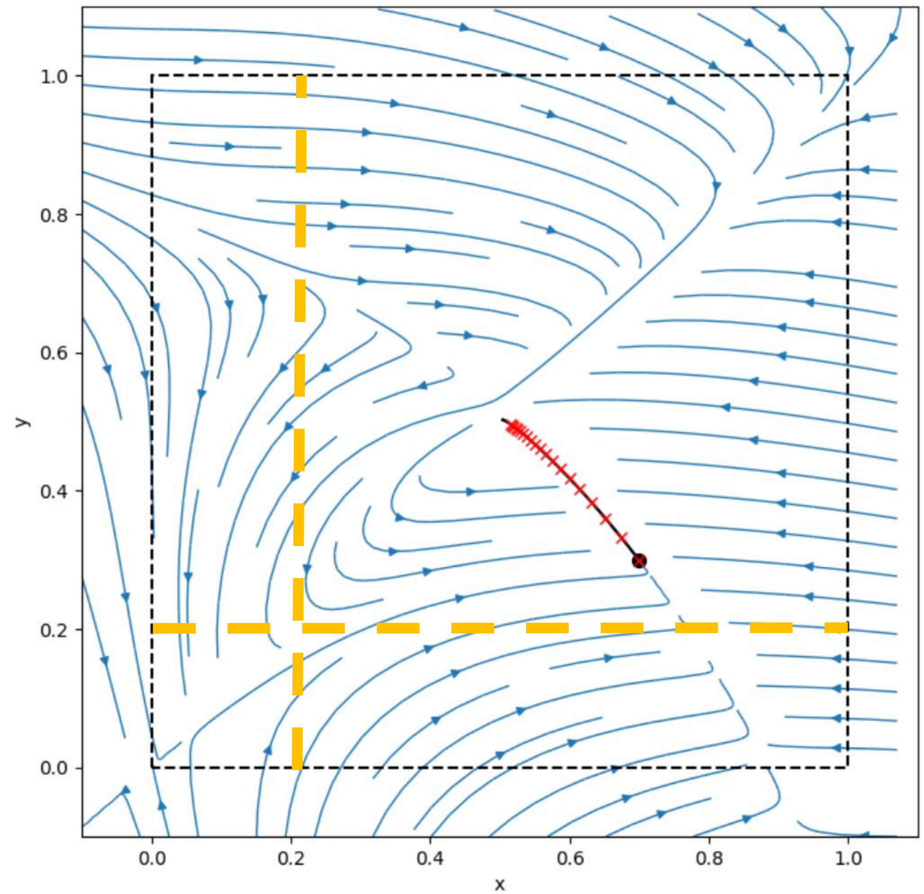
The unit square must be an invariant set!!

Learning p of degree 3

The true dynamics f (unknown)



Least squares solution subject to $p(0) = 0$, unit square invariant



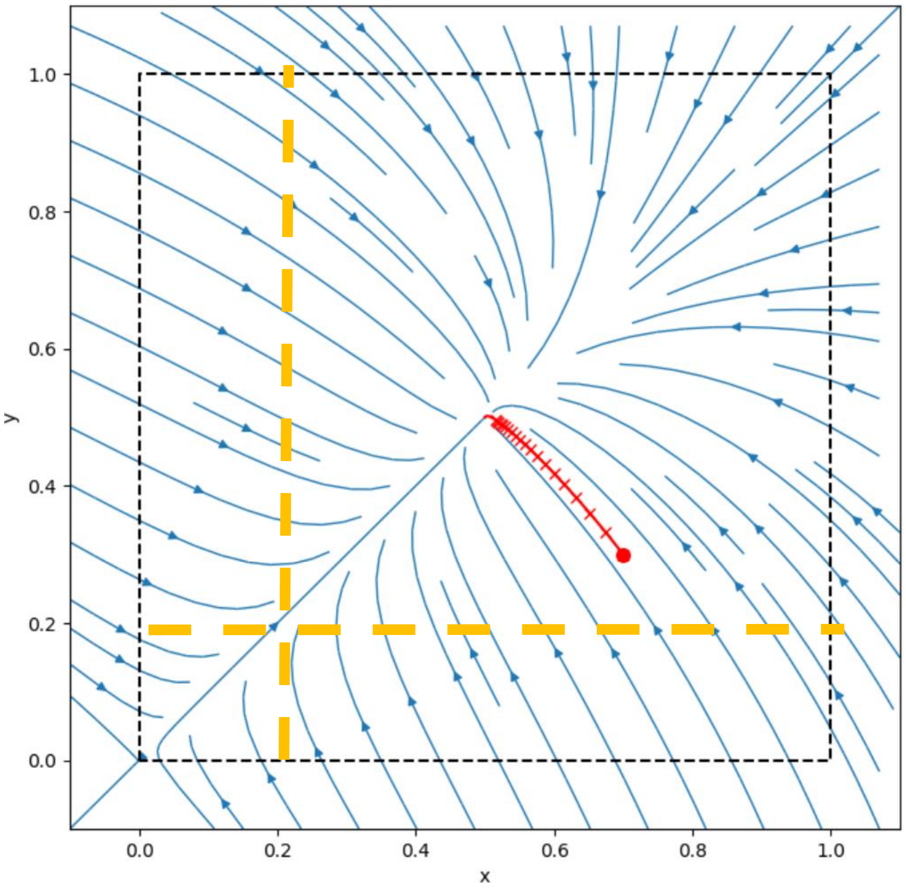
- Better, but not perfect. What other side information can you think of?

More infected females should imply higher infection rate for males! (and vice versa)

PRINCETON UNIVERSITY ORFE

Side information: directional monotonicity

The true dynamics f (unknown)



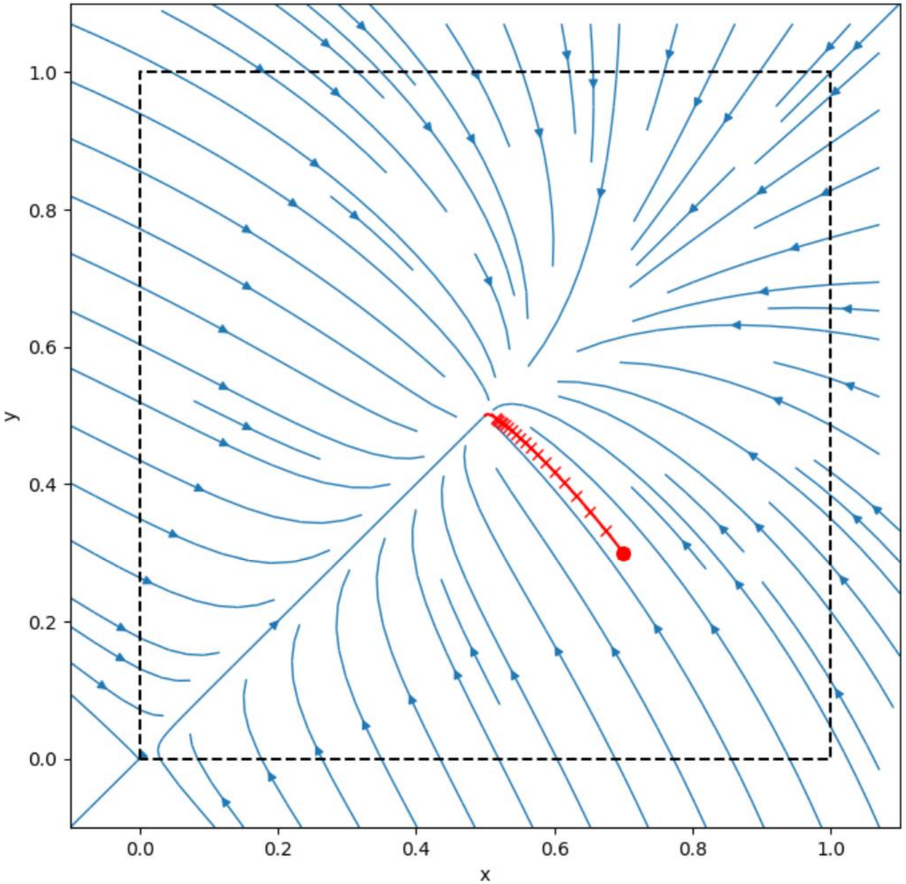
$$\frac{\partial f_1(x,y)}{\partial y} \geq 0, \forall (x,y) \in [0,1]^2$$

$$\frac{\partial f_2(x,y)}{\partial x} \geq 0, \forall (x,y) \in [0,1]^2$$

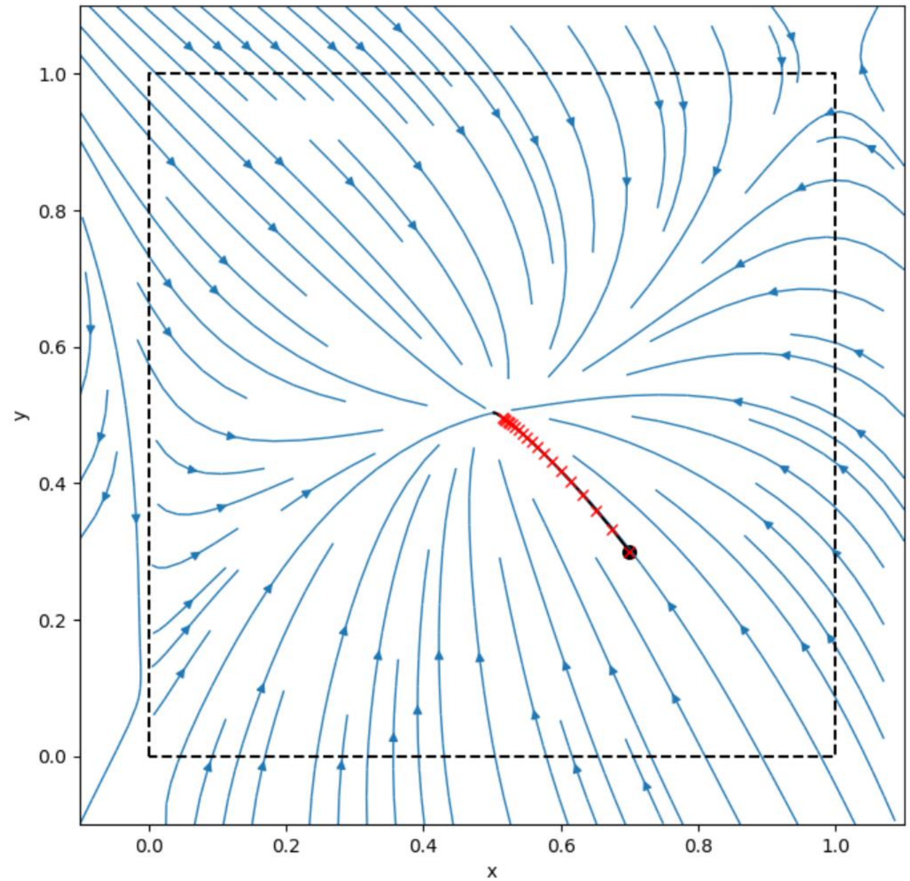
We want p to satisfy the same constraints!

Learning p of degree 3

The true dynamics f (unknown)



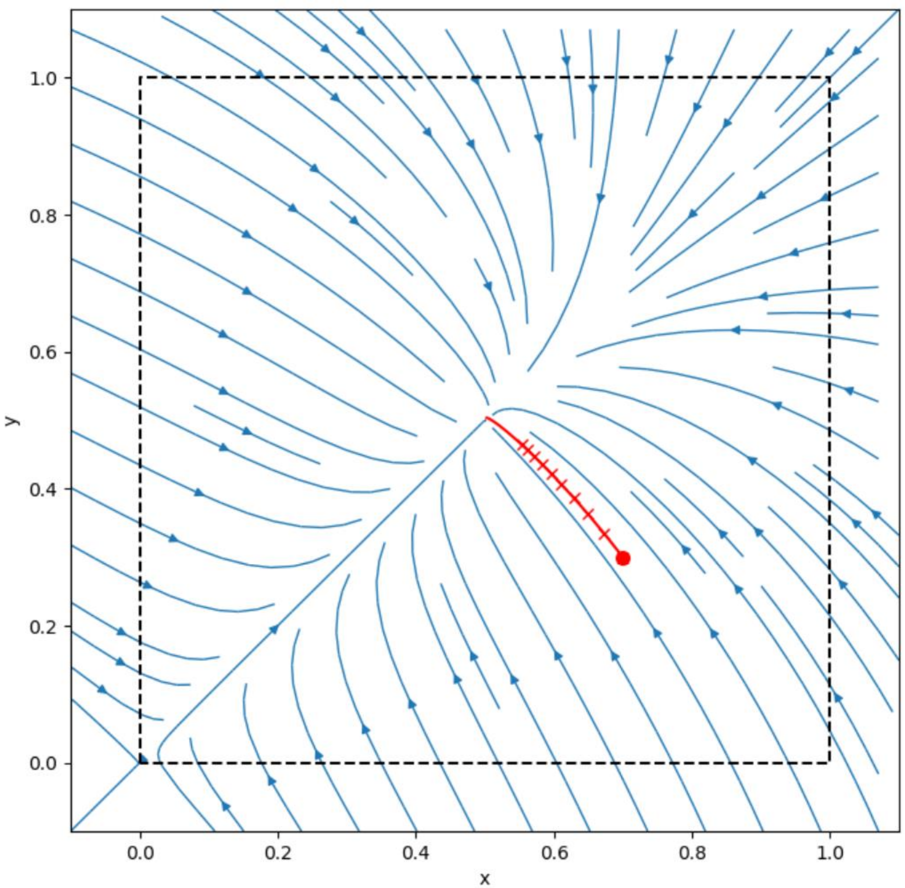
Least squares solution subject to $p(0) = 0$, unit square invariant, directional monotonicity



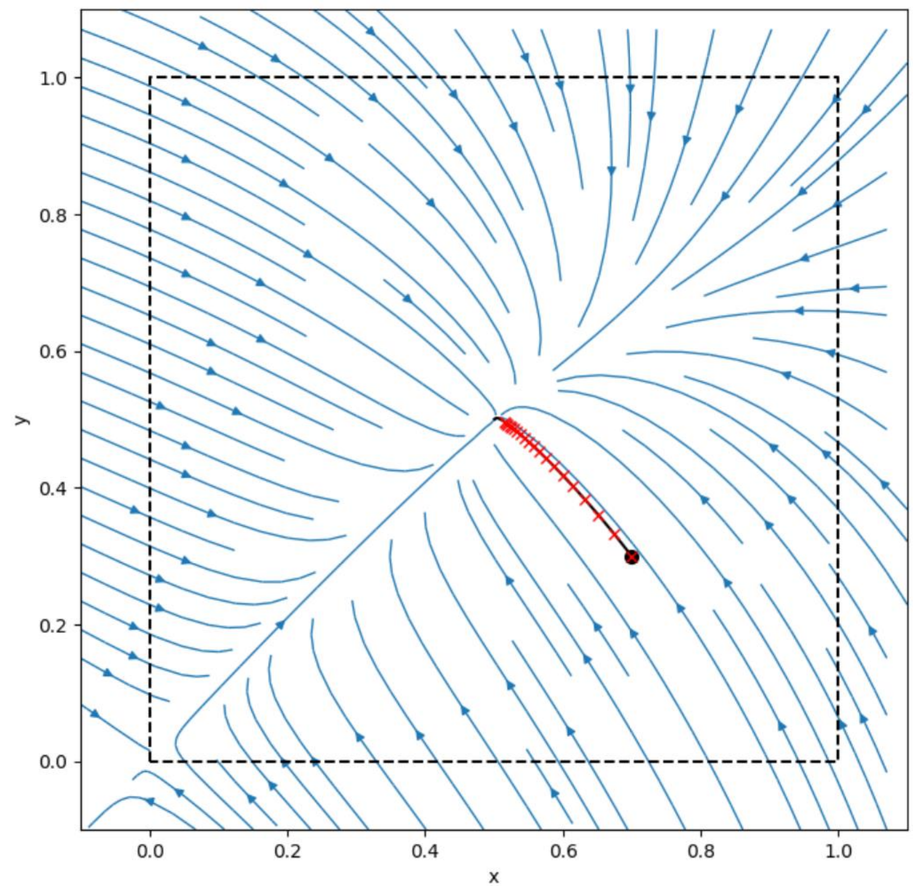
- Now we are getting the qualitative behavior correct everywhere!

Let's learn p of degree 2

The true dynamics f (unknown)



Least squares solution subject to $p(0) = 0$, unit square invariant, directional monotonicity

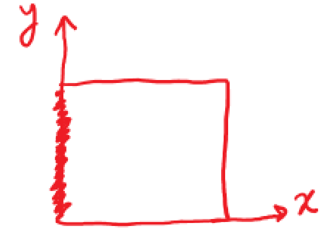


- p is pretty much dead on everywhere even though it was trained on a single trajectory!

The SDP that is being solved in the background

$$\min_{\substack{P=\begin{pmatrix}P_1\\P_2\end{pmatrix},\ \deg(P)\leq 3\\ \sigma_0,\sigma_1,\ \deg(\sigma_i)\leq 2\\ \hat\sigma_0,\hat\sigma_1,\hat\sigma_2\\ \deg(\hat\sigma_0)\leq 2}} \quad \sum_{i=1}^{20} \left(p(x^i,y^i) - \hat{f}(x^i,y^i) \right)^2$$

s.t. $P_1(0,0)=0, \quad P_2(0,0)=0$



$$\left.\begin{aligned} &P_1(0,y) = y\,\sigma_0(y) + (1-y)\,\sigma_1(y) \\ &\sigma_0,\ \sigma_1 \ \text{SOS} \end{aligned}\right\} \Rightarrow \left[x=0,\ 0\leq y\leq 1 \Rightarrow \dot{x} \geq 0\right]$$

(+ three similar constraints)

$$\left.\begin{aligned} &\frac{\partial P_1}{\partial y}(x,y) = \hat\sigma_0(x,y) + \hat\sigma_1(1-x)x + \hat\sigma_2(1-y)y \\ &\hat\sigma_0 \ \text{SOS},\ \hat\sigma_1 \geq 0,\ \hat\sigma_2 \geq 0 \end{aligned}\right\}$$

$$\Rightarrow \left[0\leq x\leq 1,\ 0\leq y\leq 1 \Rightarrow \frac{\partial P_1}{\partial y}(x,y) \geq 0\right]$$

(similarly for $\frac{\partial P_2}{\partial x}(x,y)$)

Output of SDP solver:

p1=0.2681*x^3 - 0.0361*x^2*y - 0.095*x*y^2 + 0.1409*y^3 - 0.4399*x^2 + 0.0956*x*y - 0.0805*y^2 + 0.1232*x + 0.0201*y

p2=0.1188*x^3 + 0.2606*x^2*y + 0.2070*x*y^2 + 0.0005*y^3 - 0.3037*x^2 - 0.4809*x*y -0.099*y^2+ 0.2794*x+0.01689*y

Existence of constrained polynomial dynamics close to f

Thm [AAA, El Khadir]. For any continuously differentiable vector field $f: \mathbb{R}^n \to \mathbb{R}^n$, any $T > 0, \epsilon > 0$, and any compact set $\Omega \subseteq \mathbb{R}^n$,

there exists a polynomial vector field $p: \mathbb{R}^n \to \mathbb{R}^n$ such that

1) **trajectories of $\boldsymbol{f}$ and $\boldsymbol{p}$ starting from any initial conditions $\boldsymbol{x_0} \in \boldsymbol{\Omega}$ remain within $\boldsymbol{\epsilon}$ for all time $\boldsymbol{t} \in [\boldsymbol{0}, \boldsymbol{T}]$** (as long as they stay in Ω),

2) **$\boldsymbol{p}$ satisfies any combination of the following constraints if $\boldsymbol{f}$ does:**

a. equilibria at a given finite set of points $(p(v_i) = 0)$,
b. invariance of a basic semialgebraic set $B = \{x \in \mathbb{R}^n | g_i(x) \geq 0\}$, where each g_i is concave (assumption can be relaxed),
c. directional monotonicity on a compact set $\left(\frac{\partial p_i(x)}{\partial x_j} \geq 0, \forall x \in C\right)$,
d. nonnegativity on a compact set $(p_i(x) \geq 0, \forall x \in D)$.

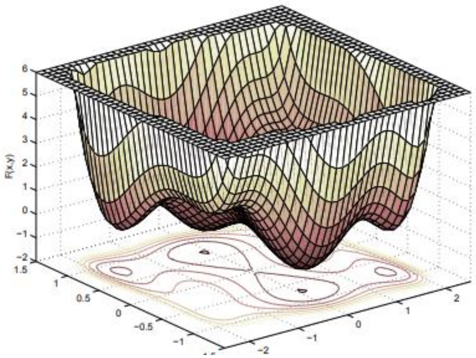
Moreover, all such properties of p come with an **SOS certificate**.

Recap: "See an inequality? Think SOS!"

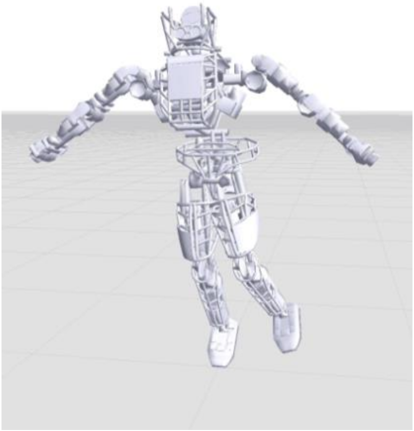
Is $p(x) \geq 0$ on $\{g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$?

Automated SOS-based proofs via SDP!

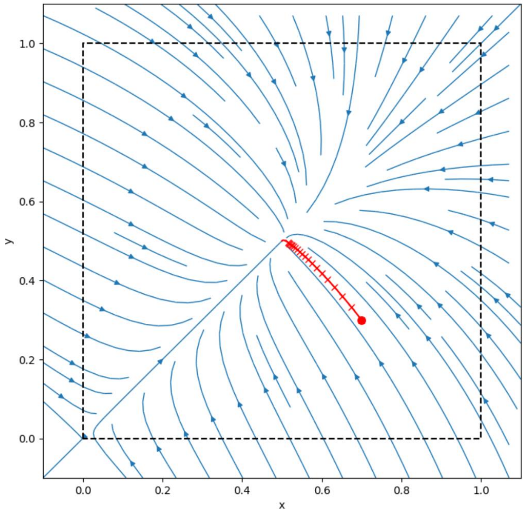
Many applications!



Optimization

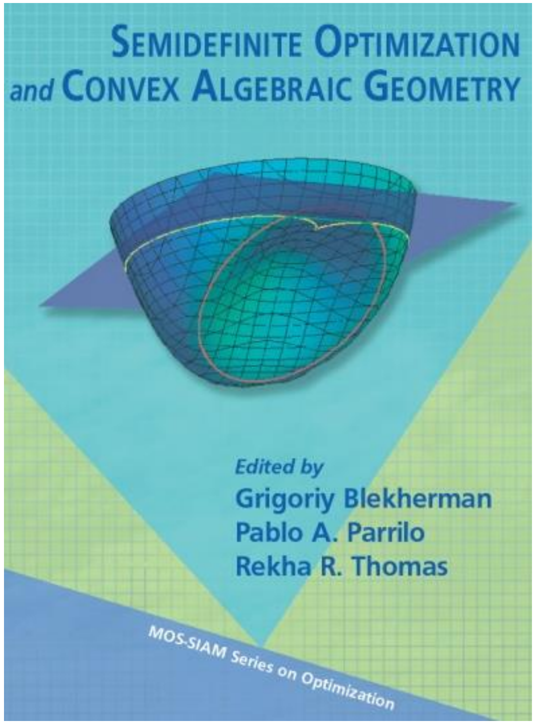


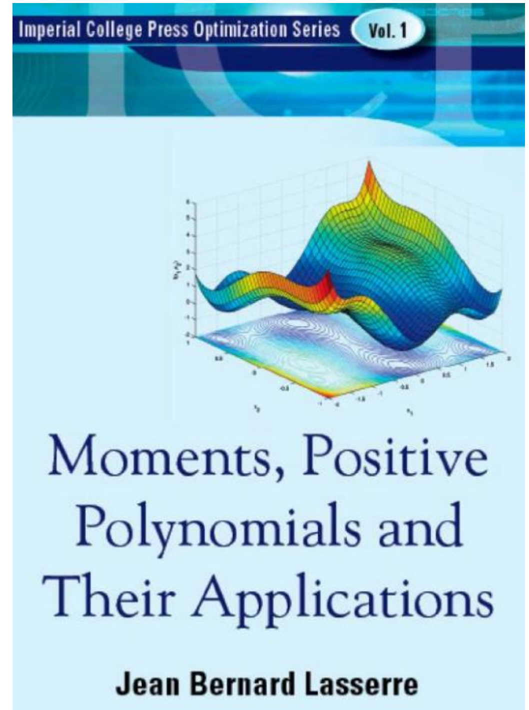
Control



Learning

Want to learn more?





SUMS OF SQUARES, MOMENT MATRICES AND OPTIMIZATION OVER POLYNOMIALS

MONIQUE LAURENT*

Applications of sums of squares

Georgina Hall