Sum of squares:
a concise introduction

Pablo A. Parrilo

LIDS - EECS - MIT

`www.mit.edu/~parrilo`

Massachusetts
Institute of
Technology

Outline

- Polynomial nonnegativity and sums of squares.
- Semidefinite programming and SOS programs.
- Certicates of infeasibility and the Positivstellensatz.
- Finding P-satz certificates using SOS/SDP.
- Application examples
- Exploiting structure: sparsity, ideals, groups and symmetries.

Nonnegativity of polynomials

How to check if a given $F(x_1, \ldots, x_n)$ is globally nonnegative?

$$F(x_1, x_2, \ldots, x_n) \geq 0, \quad \forall x \in \mathbb{R}^n$$

- For $d = 2$, easy (check eigenvalues). What happens in general?
- It is decidable, but *NP-hard* when $d \geq 4$.
- Possible approaches: Decision algebra, Tarski-Seidenberg, quantifier elimination, etc. Very powerful, but bad complexity properties.
- *Lots* of applications.
- Want “low” complexity, at the cost of possibly being conservative.

A sufficient condition: SOS

"Simple" sufficient condition: a sum of squares (SOS) decomposition:

$$F(x) = \sum_i f_i^2(x)$$

If $F(x)$ can be written as above, for some polynomials f_i, then $F(x) \geq 0$.

A purely syntactic, easily verifiable certificate.

Is this condition conservative? Can we quantify this?

- In some cases (e.g. univariate), it is exact. Full classification (Hilbert).
- Explicit counterexamples (e.g., Motzkin, Reznick, etc.)

Can we compute it efficiently?

- Yes, using semidefinite programming.

SOS and Hilbert's 17th problem

Classically, PSD=SOS for quadratics, or univariate polynomials.

Hilbert showed in 1888 that this is also true for bivariate quartics and (nonconstructively) false in all other cases.

He then asked, in 1900, as part of his famous list of 23 problems:

- Is it possible to write every psd form as $P(x) = \sum_i f_i^2(x)$, where the f_i are *rational functions*, ie. quotients of forms?
- Equivalently, does it always exist a pd $Q(x)$, such that $P(x)Q^2(x)$ is a sum of squares?

Solved by Artin ("yes!") in 1927. But, how to pick $Q(x)$?

Pólya (1928): If $F(x)$ is *pd and even*, can take $Q(x) = (\sum_i x_i^2)^r$, for r big enough.

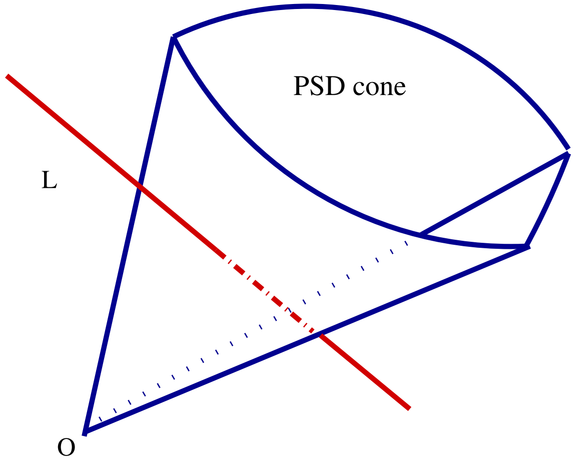
Recently (Blekherman 2006), sharp asymptotic estimates:
Vol(PSD) $= O(n^{-1/2})$, Vol(SOS) $= O(n^{-d/2})$.

Semidefinite programming - background

- A semidefinite program:

$$\{X \in \mathcal{S}^n \;:\; \mathrm{Tr} A_i X = b_i, \qquad X \succeq 0\}$$

  where $A_i \in \mathcal{S}^n$ are given symmetric matrices.



- The intersection of an affine subspace and the self-dual convex cone of positive semidefinite matrices.
- Convex finite dimensional optimization problem.
- A broad generalization of linear programming. Nice duality theory.
- Essentially, solvable in polynomial time (interior point, etc.).
- *Many, many* applications.

Checking the SOS condition

Basic method, the “Gram matrix” (Shor 87, Choi-Lam-Reznick 95, Powers-Wörmann 98, . . .)

Given $F(x)$, degree $2d$. Let z be a suitably chosen vector of monomials (in the dense case, all $\binom{n+d}{d}$ monomials of degree $\leq d$).

Then, F is SOS iff:

$$F(x) = z^T Q z, \qquad Q \succeq 0$$

- Comparing terms, obtain linear equations for the elements of Q.
- Can be solved as a semidefinite program (with equality constraints).
- Factorize $Q = L^T L$. SOS terms given by $f_i = (Lz)_i$.

SOS Example

$$\begin{aligned} F(x,y) &= 2x^4 + 5y^4 - x^2y^2 + 2x^3y \\ &= \begin{bmatrix} x^2 \\ y^2 \\ xy \end{bmatrix}^T \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix} \begin{bmatrix} x^2 \\ y^2 \\ xy \end{bmatrix} \\ &= q_{11}x^4 + q_{22}y^4 + (q_{33} + 2q_{12})x^2y^2 + 2q_{13}x^3y + 2q_{23}xy^3 \end{aligned}$$

(notice: Q is not unique – important!)

Since Q must be PSD, this is an SDP. Solving, we obtain:

$$Q = \begin{bmatrix} 2 & -3 & 1 \\ -3 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix} = L^T L, \qquad L = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & -3 & 1 \\ 0 & 1 & 3 \end{bmatrix}$$

And therefore

$$F(x,y) = \tfrac{1}{2}(2x^2 - 3y^2 + xy)^2 + \tfrac{1}{2}(y^2 + 3xy)^2$$

(notice: number of squares is the rank of Q)

Sum of squares programs

A SOS program is an optimization problem with SOS constraints:

$$\begin{array}{rll} \min_{u_i} & c_1 u_1 + \cdots + c_n u_n & \\ \text{s.t} & P_i(x,u) := A_{i0}(x) + A_{i1}(x) u_1 + \cdots + A_{in}(x) u_n & \text{are SOS} \end{array}$$

- Convex finite dimensional optimization problems.
- Many problems have very natural formulations (or relaxations) as SOS programs.
- Can convert to SDPs – if necessary.
- Many applications, besides optimization: control and dynamical systems, geometric theorem proving, quantum information theory, etc.

(Aside: Can we do this black-box? Even for a single poly?)

Example: Spherical codes

A *spherical code* is a set of N points $v_i \in S^{d-1}$, with $\measuredangle(v_i, v_j) \geq \theta$.

The "LP/Delsarte-Goethals-Seidel/Kabatiansky-Levenshtein" upper bound on the size of a d-dimensional spherical code of minimum angle θ:

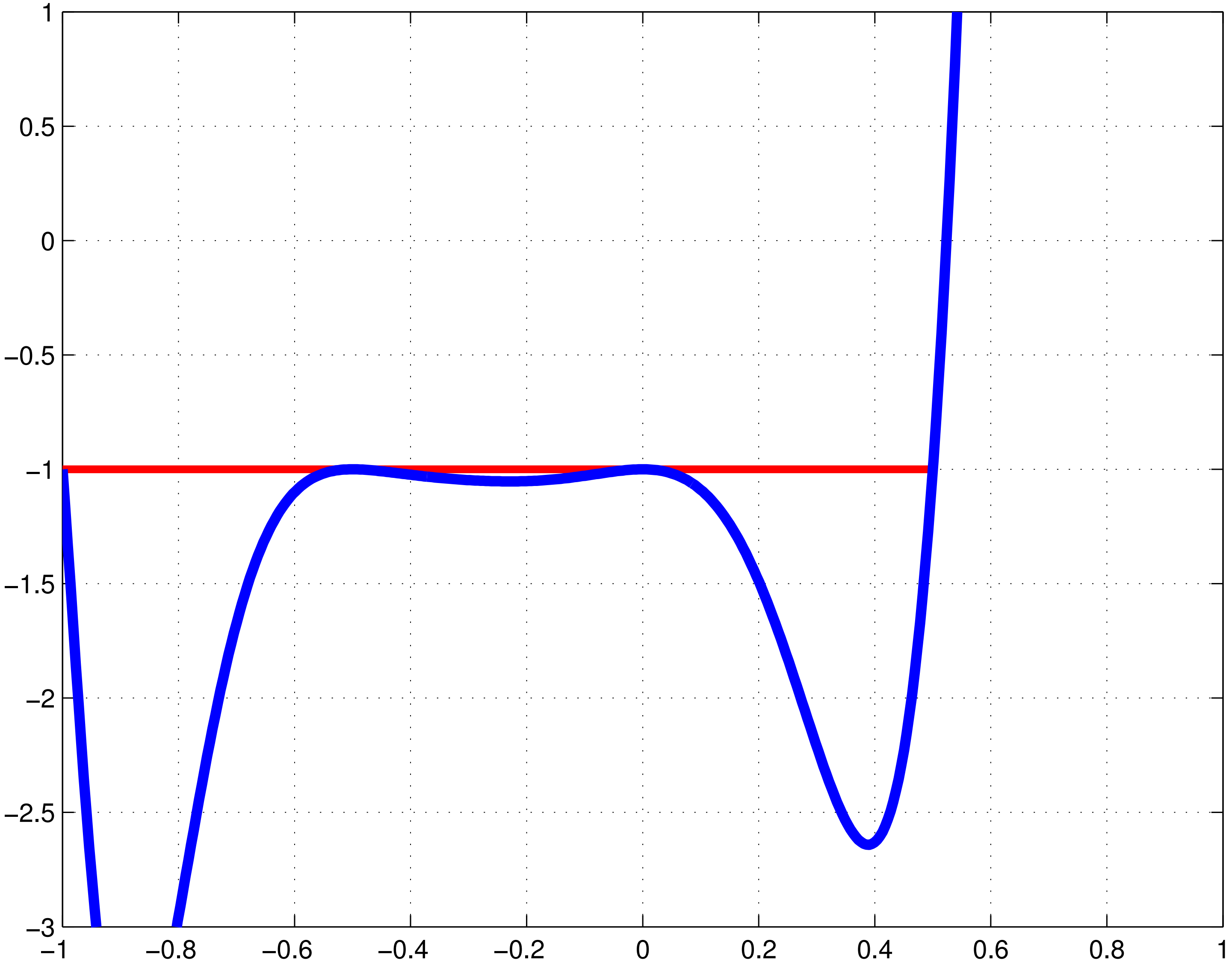
$$\min \sum_{k=1}^{m} c_k \qquad \text{s.t.} \quad \begin{cases} \sum_{k=1}^{m} c_k \, P_k(t) \; \leq \; -1 & \forall t \in [-1, \cos\theta] \\ c_k \; \geq \; 0 \end{cases}$$

where the $P_k(t)$ are Gegenbauer polynomials (of parameter $(d-3)/2$).

Can formulate this as an SOS problem.
Instead of discretizing over t, solve directly as an SDP!

(Similarly for hypercube, Grassmannian, 2-point homogeneous spaces, etc. Conceptually, indep. set in large graph + Lovász ϑ + symmetry reduction).

E.g., for $d = 8$, $m = 6$, $\theta = \frac{\pi}{3}$, SDP bound yields the exact value $\mathbf{240}$.



(A matching lower bound, E_8 lattice code). **Q:** asymptotics?

Polynomial systems over the reals

- When does a system of equations and inequalities have real solutions?

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

- A remarkable answer: the Positivstellensatz.
- A fundamental theorem in real algebraic geometry, due to Stengle.
- A common generalization of Hilbert's Nullstellensatz and LP duality.
- Guarantees the existence of infeasibility certificates for real solutions of systems of polynomial equations.
- Sums of squares are a fundamental ingredient.

How does it work?

Linear programming duality

Certificates nonexistence of real solutions of linear equations.

$$\left\{ \begin{array}{rl} Ax + b & \geq 0 \\ Cx + d & = 0 \end{array} \right\} = \emptyset \quad \Longleftrightarrow \quad \exists\, \lambda, \nu \ \text{s.t.} \ \left\{ \begin{array}{rl} \lambda^T A + \nu^T C & = 0 \\ \lambda^T b + \nu^T d & = -1 \\ \lambda & \geq 0 \end{array} \right.$$

- Finding certificates is also a linear programming problem.
- Also known as Farkas' lemma.
- Primal and dual are polynomial time solvable.
- Relies on convexity.

Well known, but there are more...

Hilbert's Nullstellensatz

Certificates nonexistence of complex solutions of polynomial equations.

$$\{z \in \mathbb{C}^n |\ f_i(z) = 0\} = \emptyset \qquad \Longleftrightarrow \qquad \begin{array}{c} 1 \in \textbf{ideal}(f_i) \\ \text{or} \\ \exists q_i(z) \text{ s.t. } \sum_i f_i(z) q_i(z) = 1 \end{array}$$

- Cornerstone of algebraic geometry, establishes a correspondence between geometric ideas and algebraic objects.

affine varieties ⇔ polynomial ideals

- "Canonical" NP-complete problem in the real model of computation.
- If q_i have fixed degree, can solve using linear algebra.

How to generalize this?

Degree \ Field	Complex	Real
Linear	Range/Kernel	LP duality
Polynomial	Nullstellensatz	?????

Can we get the best of both worlds?

General *polynomial* equations, as in the Nullstellensatz.

And *real* solutions, so we can handle inequalities?

HOW?

Positivstellensatz

Given $\{x \in \mathbb{R}^n \mid f_i(x) \geq 0, \quad h_i(x) = 0\}$. Define:

$$\mathbf{cone}(f_i) = \sum s_i \cdot (\prod_j f_j), \qquad \mathbf{ideal}(h_i) = \sum t_i \cdot h_i,$$

where the $s_i, t_i \in \mathbb{R}[x]$ and the s_i are sums of squares.

To prove infeasibility, find $f \in \mathbf{cone}(f_i), h \in \mathbf{ideal}(h_i)$ such that

$$f + h = -1.$$

- A fundamental theorem in real algebraic geometry (Stengle 1974).
- Provides *infeasibility certificates*, generalizes Lagrangian duality.
- Unless NP=co-NP, the certificates cannot *always* be polynomially sized.

P-satz and SDP

Certificates for real solutions of systems of polynomial equations!

$$\left\{ \begin{array}{rcl} x & \in & \mathbb{R}^n \\ f_i(x) & \geq & 0 \\ h_i(x) & = & 0 \end{array} \right\} = \emptyset \qquad \Longleftrightarrow \qquad \exists f, h \quad \left\{ \begin{array}{l} f + h = -1 \\ f \in \mathbf{cone}(f_i) \\ h \in \mathbf{ideal}(h_i) \end{array} \right.$$

- The condition is convex in the unknowns $s_i(x), t_i(x)$.
- Thus, the set of emptiness proofs is convex!
- For bounded degree, can find certificates by solving SDPs!
- A complete *SDP hierarchy*, given by certificate degree.
- Tons of applications:
  optimization, dynamical systems, quantum mechanics...

The dual view: pseudoexpectations

If S is nonempty, any probability distribution μ on it satisfies

$$\begin{aligned} \mathbf{E}_\mu(q) &\geq 0 \qquad \forall q \in \mathbf{cone}(f_i) \\ \mathbf{E}_\mu(q) &= 0 \qquad \forall q \in \mathbf{ideal}(h_i) \end{aligned}$$

Dual variables are linear functionals $\tilde{\mathbf{E}}_\mu : \mathbb{R}[x]_{2k} \to \mathbb{R}$

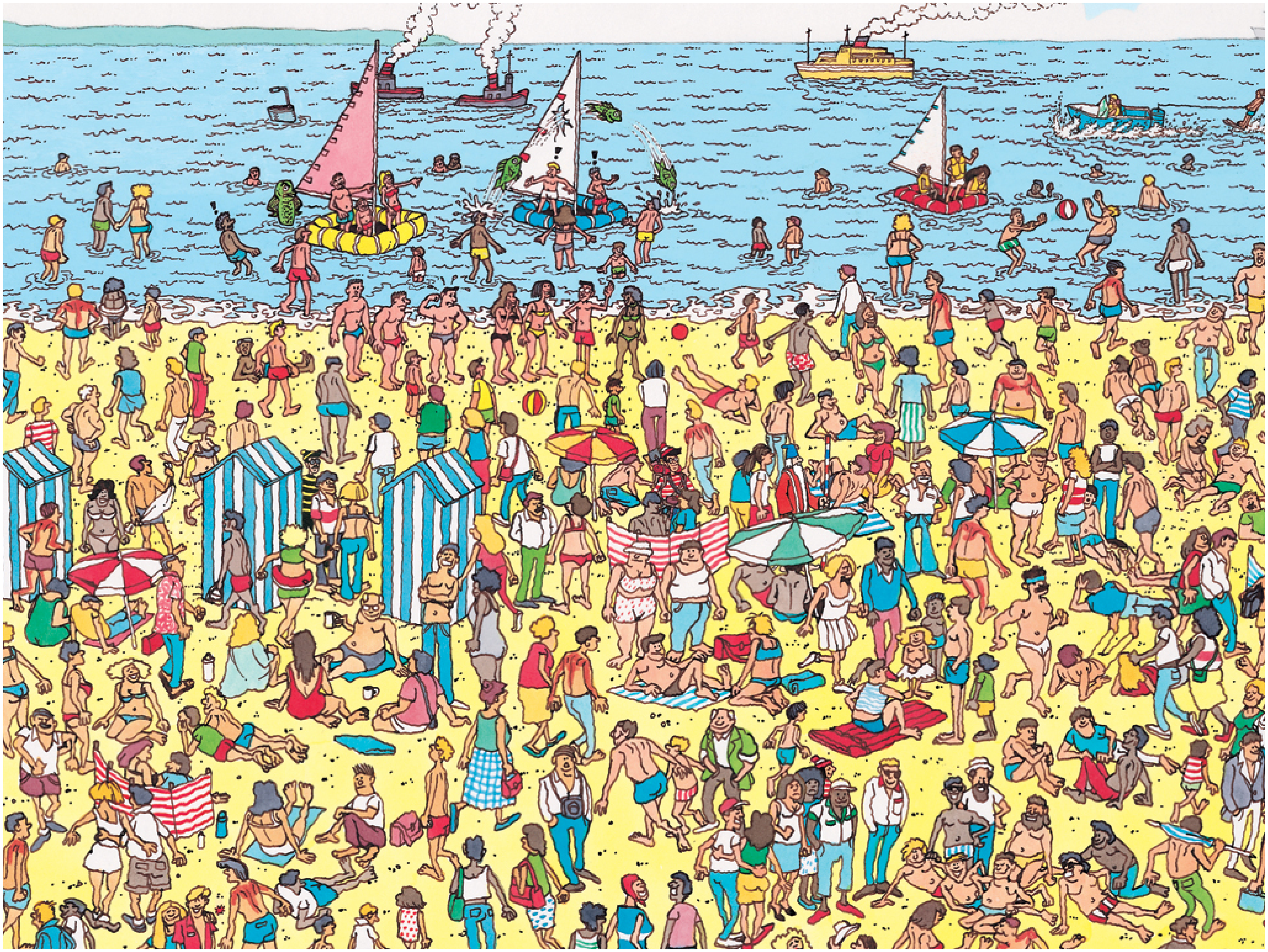
Connections to classical moment problem (“moment sequences”, “Riesz functional”, “pseudoexpectation”, . . .)

Test: If S is nonempty, there always exist $\tilde{\mathbf{E}}_\mu$ such that

$$\begin{aligned} \tilde{\mathbf{E}}_\mu(q) &\geq 0 \qquad \forall q \in \mathbf{cone}(f_i), \quad \mathbf{sdeg}(q) \leq 2k \\ \tilde{\mathbf{E}}_\mu(q) &= 0 \qquad \forall q \in \mathbf{ideal}(h_i), \quad \mathbf{sdeg}(q) \leq 2k. \end{aligned}$$

A (possibly fake) solution, that passes low-degree tests.

Where's Waldo?



Where's Waldo? (behind a frosted glass)

Where’s Waldo? (behind a very frosted glass)

P-satz and SDP

The "true" problem:

$$\left\{\begin{array}{rcl} x & \in & \mathbb{R}^n \\ f_i(x) & \geq & 0 \\ h_i(x) & = & 0 \end{array}\right\} = \emptyset \qquad \Longleftrightarrow \qquad \exists f, h \quad \left\{\begin{array}{l} f + h = -1 \\ f \in \mathbf{cone}(f_i) \\ h \in \mathbf{ideal}(h_i) \end{array}\right.$$

"Bounded resources/degree" version (SOS hierarchy):

$$\left\{\begin{array}{rcl} \tilde{\mathbf{E}}_\mu & \in & \mathbb{R}[x]_{2k} \\ \tilde{\mathbf{E}}_\mu[q] & \geq & 0, \quad \forall q \in \mathbf{cone}(f_i) \\ \tilde{\mathbf{E}}_\mu[q] & = & 0, \quad \forall q \in \mathbf{ideal}(f_i) \\ & & \mathbf{sdeg}(q) \leq 2k \end{array}\right\} = \emptyset \quad \Longleftrightarrow \quad \exists f, h \quad \left\{\begin{array}{l} f + h = -1 \\ f \in \mathbf{cone}(f_i) \\ h \in \mathbf{ideal}(h_i) \\ \mathbf{sdeg}(f), \mathbf{sdeg}(h) \leq 2k \end{array}\right.$$

Slightly different versions (e.g., Lasserre), depending on assumptions or representation theorems (Putinar, Schmudgen, Pólya, . . .).

Example: Grigoriev's knapsack

Mother of all (TCS) examples: infeasible system, but no short SOS proof.

Consider the 0/1 system with $0 \leq r \leq n/2$ (infeasible for noninteger r):

$$\sum_{i=1}^{n} x_i = r, \qquad x_i^2 = x_i, \quad i = 1, \ldots, n.$$

"Obvious" pseudo-expectations (dual variables) on squarefree monomials:

$$\tilde{\mathbf{E}}_\mu[x_S] := \binom{r}{|S|} / \binom{n}{|S|}$$

Can show this fools all SOS tests of degree $\Omega(r)$.
(e.g., explicit diagonalization, Johnson scheme, Krawtchouk, ...)

SOS-proofs of many inequalities

- Cauchy-Schwarz as polynomial inequality (Lagrange's identity)

$$\|\mathbf{x}\|^2\|\mathbf{y}\|^2 - \langle \mathbf{x}, \mathbf{y}\rangle^2 = \sum_{1\le i<j\le n} (x_i y_j - x_j y_i)^2 \ge 0.$$

- Generalized AGM follows from the SOS inequality (Hurwitz):

$$x_1^{2d} + x_2^{2d} + \cdots + x_{2d}^{2d} - (2d) x_1 x_2 \cdots x_{2k} \text{ is SOS.}$$

- Cauchy-Schwarz for pseudoexpectations:

$$\left(\tilde{\mathbf{E}}_\mu[pq]\right)^2 \le \left(\tilde{\mathbf{E}}_\mu[p^2]\right)\left(\tilde{\mathbf{E}}_\mu[q^2]\right).$$

Pf:

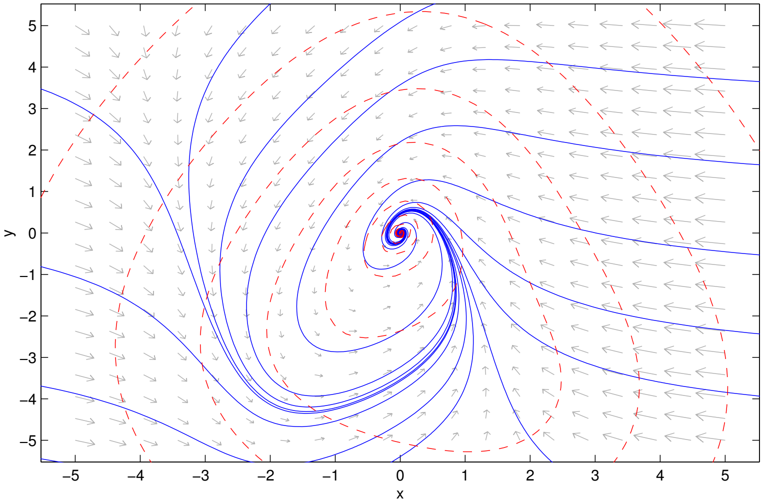
$$\forall \alpha, \beta \in \mathbb{R} \quad \tilde{\mathbf{E}}_\mu[(\alpha p + \beta q)^2] \ge 0 \quad \Leftrightarrow \quad \begin{bmatrix} \tilde{\mathbf{E}}_\mu[p^2] & \tilde{\mathbf{E}}_\mu[pq] \\ \tilde{\mathbf{E}}_\mu[pq] & \tilde{\mathbf{E}}_\mu[q^2] \end{bmatrix} \succeq 0.$$

Why is this a good approach?

Properties we desire from numerical methods:

- "Easy" problems should remain easy. Special cases should give sensible methods for simplified problems (e.g., spectral methods)
- Coordinate and representation invariance. The solution time should not depend on the coordinates chosen, but rather on the underlying geometric object. One (big) exception: sparse problems.
- Certification of solutions. Shouldn't need to trust software.
- "Practical" problems are (usually) nondegenerate, and have structure. We can exploit many kinds of structure.
- To find solutions, need rounding methods!
- A direct generalization of the best available techniques.

Stability of dynamical systems



- Given a system of ODEs

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0$$

- Want to prove stability, i.e., solutions converge to zero for all initial conditions

- To prove this, need to find an energy-like *Lyapunov function*:

$$V(x) \geq 0, \qquad \qquad \dot{V}(x) := \left(\frac{\partial V}{\partial x}\right)^T f(x) \leq 0$$

- With an affine family of candidate polynomial V, $\dot{V}$ is also affine.

- Instead of checking nonnegativity, use a SOS condition.

- Many variations: uncertain parameters, time delays, PDEs, etc.

Deciding quantum entanglement

A bipartite mixed quantum state ρ is *separable* (not *entangled*) if

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \otimes |\phi_i\rangle\langle\phi_i| \qquad \sum p_i = 1,$$

for some ψ_i, ϕ_i.

(Essentially, convex hull of tensor product of rank-one matrices).

Given ρ, how to decide if it is entangled?

- Essentially, equivalent to nonnegativity of biquadratic polys
- Yields a hierarchy of SDP-based entanglement witnesses
- The first level, corresponds to a well-known criterion (PPT).
- Related problem: optimizing over set of separable states.

(e.g., Doherty-P.-Spedalieri, Brandão-Christandl-Yard, Barak-Khotari-Steurer, . . .)

Exploiting structure

What algebraic properties of the polynomial system yield efficient computation?

- *Sparsity:* few nonzero coefficients.
  - Newton polytopes techniques
  - Complexity does not depend on the degree.
- *Symmetries:* invariance under a transformation group.
  - Frequent in practice. Enabling factor in applications.
  - Can reflect underlying physical symmetries, or modelling choices.
  - Representation theory and invariant-theoretic techniques.
- *Ideal structure:* Equality constraints.
  - SOS on *quotient rings*.
  - Compute in the coordinate ring. Quotient bases (Gröbner).
  - E.g., interesting varieties (Veronese, Segre, Grassmannian, etc.)

Symmetry reduction

In practice, many problems are invariant under a group of transformations.

$$p(x) = p(tx), \qquad \forall t \in T$$

where $T \subseteq GL(\mathbb{R}^n)$ is a matrix group.

- Ex: $\min x^4 + y^4 + z^4 - 4xyz + x + y + z$.

  Invariant under permutations of x, y, z.

- Ex: Grigoriev's knapsack

- Ex: Nonnegativity of *even* forms (copositivity).

What are the geometric, algebraic, and computational implications?

Exploitation of symmetries is an enabling factor in applications.

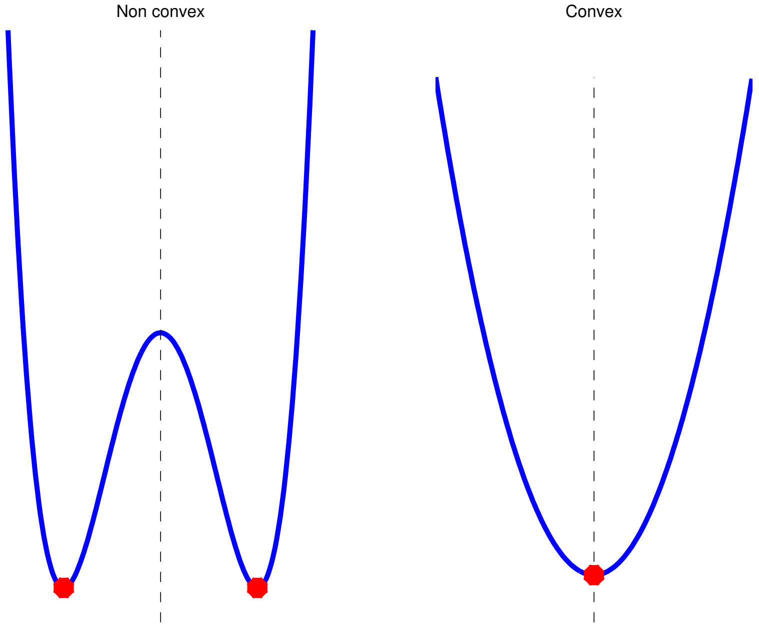
How to do this in the SOS/SDP framework?

Symmetry and convexity

Key property of symmetric *convex* sets: the "group average" $\frac{1}{|G|}\sum_{g\in G}\sigma(g)x$ always belongs to the set.

So, in convex optimization can always restrict to the fixed-point subspace

$$\{x|\sigma(g)x = x, \quad \forall g \in G\}.$$



Instead of looking for solutions in the original space, use the orbit (quotient) space.

In SDP, the restriction to the fixed-point subspace takes the form:

$$\sigma(g)M = M \qquad \Longrightarrow \qquad \rho(g)M - M\rho(g) = 0, \quad \forall g \in G. \tag{1}$$

Schur's lemma of representation theory *exactly characterizes* the matrices that commute with a group action.

Decomposing the problem

Using Schur's lemma, every group representation decomposes as a direct sum of N irreducible representations:

$$\rho = m_1\vartheta_1 \oplus m_2\vartheta_2 \oplus \cdots \oplus m_N\vartheta_N$$

where $m_1, \ldots, m_N$ are the multiplicities. Therefore, an isotypic decomposition:

$$\mathbb{C}^n = V_1 \oplus \cdots \oplus V_N, \quad V_i = V_{i1} \oplus \cdots \oplus V_{in_i}.$$

In the symmetry-adapted basis, matrix M in (1) has a block diagonal form:

$$M = (I_{m_1} \otimes M_1) \oplus \ldots \oplus (I_{m_N} \otimes M_N)$$

- Not only the SDP block-diagonalizes, but also many blocks are identical!
- Smaller, coupled problems.
- Instead of checking if a big matrix is PSD, use one M_i per block!

SOS over everything...

We can interpret other features (equality constraints, symmetries) as doing SOS over different algebraic structures. The algebraic language is *essential* to exploit problem structure:

Standard	Equality constraints	Symmetries
polynomial ring $\mathbb{R}[x]$	quotient ring $\mathbb{R}[x]/I$	invariant ring $\mathbb{R}[x]^G$
monomials (deg $\leq k$)	*standard* monomials	isotypic components
$\frac{1}{(1-\lambda)^n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} \cdot \lambda^k$	Hilbert series	Molien series
	Finite convergence for zero dimensional ideals	Block diagonalization

The different techniques are *mutually compatible*.

SOS + symmetry reduction $\to$ Flag algebras

What happens for (symmetric) fixed-degree polynomials, in a very large (or infinite) number of variables?

(e.g., “Symmetric sums of squares over k-subset hypercubes,” by A. Raymond/J. Saunderson/M. Singh/R. Thomas)

- SDP-like description that is independent of n, as $n \to \infty$.
- Very useful, for instance, to understand relationships between subgraph densities in graphs (invariant under node relabelings)

Connections to Razborov’s flag algebras, graphons / graph limits, etc.

More tools for asymptotic analysis? How to solve/analyze these infinite SDPs?

Outstanding example: what happens asymptotically with higher-degree SOS bounds (e.g., Schrijver’s) for coding/packing?

Summary

- Powerful machinery, synergy between algebra and optimization
- Classical roots, connections with many exciting areas of mathematics
- Remarkably, useful in both theory and practice
- Broad, natural generalization of earlier results.
- General construction translates into efficient, concrete results.
- Much recent progress, particularly on rounding/analysis
- Need help! Asymptotic analysis, fast algorithms, etc.
- Lots of interesting things to do!