Sum of squares techniques and polynomial optimization

Pablo A. Parrilo

Laboratory for Information and Decision Systems
Electrical Engineering and Computer Science
Massachusetts Institute of Technology
www.mit.edu/~parrilo

CDC 2016 - Las Vegas

Massachusetts Institute of Technology

Polynomial problems

We will discuss optimization and decision problems involving *multivariate polynomials*.

Usually, this means a standard optimization problem

$$\min f(x) \qquad \text{s.t.} \quad g_i(x) \leq 0,$$

where the objective and constraints are polynomial expressions.

We may also have (slightly) more complicated *quantified* formulas, and problems where the *variables* are themselves polynomials.

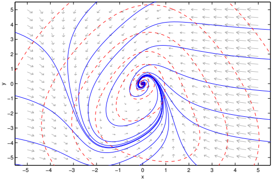
Focus on the basic ideas, emphasizing the geometric and complexity aspects. Much more is known.

Where do these problems appear?

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., that solutions converge to the origin for all initial conditions
- To prove this, need to find an energy-like *Lyapunov function*:

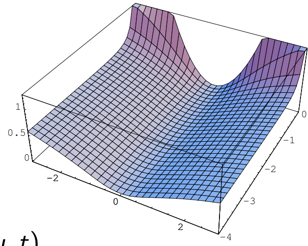
$$V(x) \geq 0, \qquad \dot{V}(x) := \left(\frac{\partial V}{\partial x}\right)^T f(x) \leq 0$$

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

Partial differential inequalities

- Solutions for linear PDIs:
  - Dissipation or Lyapunov:

$$V(x) \geq 0, \quad \left(\frac{\partial V}{\partial x}\right)^T f(x) \leq 0, \quad \forall x$$



  - Hamilton-Jacobi:

$$V(x,t) \geq 0, \quad -\frac{\partial V}{\partial t} + \mathcal{H}(x, \frac{\partial V}{\partial x}) \leq 0, \quad \forall (x,u,t)$$

- Very difficult in three or higher dimensions.
- Many approaches: approximation, discretization, level-set methods...

How to find certified solutions?
Can we obtain bounds on linear functionals of the solutions?

Motivation

Common properties:

- Can be expressed/approximated with polynomials and/or rational functions
- Include nonnegativity constraints (perhaps implicitly)
- Provably difficult (NP-complete, or worse)

These correspond to a very large class of problems:
quantified polynomial inequalities or semialgebraic problems.

Roadmap

- Motivating examples
- Optimization over polynomials
- Sum of squares programs
  - Convexity, relationships with semidefinite programming
  - Geometric interpretations
- Certificates
- Examples: extremal polynomials, joint spectral radius
- Exploiting structure: algebraic and numerical techniques.
- Perspectives, challenges, open questions

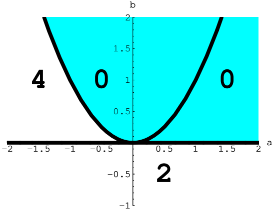
Things get complicated...

The set of nonnegative polynomials is not *basic* semialgebraic.

The set $\{(a_1, \ldots, a_n) \mid \sum_{k=1}^n a_k x^k \geq 0 \ \forall x\}$ *cannot* be described using a *finite* number of unquantified polynomial inequalities $g_i(a_1, \ldots, a_n) \geq 0$.

Ex: Consider the convex set of (a, b) for which

$$x^4 + 2ax^2 + b \geq 0 \quad \forall x \in \mathbb{R}$$



This set *cannot* be defined by $\{g_i(a, b) \geq 0\}$.

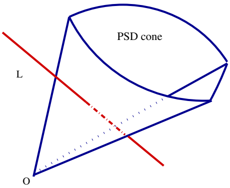
Gets worse in higher dimensions. We need either:

- Boolean set operations (unions of basic SA sets)
- Embed in higher dimensional spaces (lift and project)

Semidefinite programming (LMIs)

A broad generalization of LP to symmetric matrices

$$\min \operatorname{Tr} CX \qquad \text{s.t.} \quad X \in \mathcal{L} \cap \mathcal{S}_+^n$$



- The intersection of an affine subspace $\mathcal{L}$ and the cone of positive semidefinite matrices.
- *Lots* of applications. A true "revolution" in computational methods for engineering applications
- First applications in control theory and combinatorial optimization. Nowadays, applied everywhere.
- Convex finite dimensional optimization. Nice duality theory.
- Essentially, solvable in polynomial time (interior point, etc.)

Sum of squares

As we have seen, handling nonnegativity directly is too difficult. Instead. . .

A multivariate polynomial $p(x)$ is a sum of squares (SOS) if

$$p(x) = \sum_i q_i^2(x), \quad q_i(x) \in \mathbb{R}[x].$$

- If $p(x)$ is SOS, then clearly $p(x) \geq 0 \;\; \forall x \in \mathbb{R}^n$.
- *Convex* condition: p_1, p_2 SOS $\Rightarrow \lambda p_1 + (1-\lambda)p_2$ SOS for $0 \leq \lambda \leq 1$.
- SOS polynomials form a convex cone

For univariate or quadratic polynomials, SOS and nonnegativity are equivalent.

From LMIs to SOS

LMI optimization problems:

> affine families of *quadratic* forms, that are *nonnegative*.

Instead, for SOS we have:

> affine families of *polynomials*, that are *sums of squares*.

An SOS program is an optimization problem with SOS constraints:

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

This is a finite-dimensional, convex optimization problem.

From LMIs to SOS

LMI optimization problems:

affine families of *quadratic* forms, that are *nonnegative*.

Instead, for SOS we have:

affine families of *polynomials*, that are *sums of squares*.

An SOS program is an optimization problem with SOS constraints:

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

This is a finite-dimensional, convex optimization problem.

From LMIs to SOS

LMI optimization problems:

affine families of *quadratic* forms, that are *nonnegative*.

Instead, for SOS we have:

affine families of *polynomials*, that are *sums of squares*.

An SOS program is an optimization problem with SOS constraints:

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

This is a finite-dimensional, convex optimization problem.

SOS programs: questions

- Why not just use nonnegative polynomials?
  While convex, unfortunately it's NP-hard ;(
- And is SOS any better?
  Yes, we can solve SOS programs in polynomial time
- Aren't we losing too much then?
  In several important cases (quadratic, univariate, etc), nonnegativity and SOS is the same thing.
- And in the other cases?
  Low dimension, computations and some theory show small gap.
  Recent negative results in very high dimension, though (Blekherman)
- Isn't it a very special formulation?
  No, we can approximate *any* semialgebraic problem!
- How? And how do you solve them?

OK, I'll tell you. But first, an example!

SOS programs: questions

- Why not just use nonnegative polynomials?
  While convex, unfortunately it's NP-hard ;(
- And is SOS any better?
  Yes, we can solve SOS programs in polynomial time
- Aren't we losing too much then?
  In several important cases (quadratic, univariate, etc), nonnegativity and SOS is the same thing.
- And in the other cases?
  Low dimension, computations and some theory show small gap.
  Recent negative results in very high dimension, though (Blekherman)
- Isn't it a very special formulation?
  No, we can approximate *any* semialgebraic problem!
- How? And how do you solve them?
  OK, I'll tell you. But first, an example!

SOS programs: questions

- Why not just use nonnegative polynomials?
  While convex, unfortunately it's NP-hard ;(
- And is SOS any better?
  Yes, we can solve SOS programs in polynomial time
- Aren't we losing too much then?
  In several important cases (quadratic, univariate, etc), nonnegativity and SOS is the same thing.
- And in the other cases?
  Low dimension, computations and some theory show small gap.
  Recent negative results in very high dimension, though (Blekherman)
- Isn't it a very special formulation?
  No, we can approximate *any* semialgebraic problem!
- How? And how do you solve them?

OK, I'll tell you. But first, an example!

SOS programs: questions

- Why not just use nonnegative polynomials?
  While convex, unfortunately it's NP-hard ;(
- And is SOS any better?
  Yes, we can solve SOS programs in polynomial time
- Aren't we losing too much then?
  In several important cases (quadratic, univariate, etc), nonnegativity and SOS is the same thing.
- And in the other cases?
  Low dimension, computations and some theory show small gap.
  Recent negative results in very high dimension, though (Blekherman)
- Isn't it a very special formulation?
  No, we can approximate *any* semialgebraic problem!
- How? And how do you solve them?

OK, I'll tell you. But first, an example!

SOS programs: questions

- Why not just use nonnegative polynomials?
  While convex, unfortunately it's NP-hard ;(
- And is SOS any better?
  Yes, we can solve SOS programs in polynomial time
- Aren't we losing too much then?
  In several important cases (quadratic, univariate, etc), nonnegativity and SOS is the same thing.
- And in the other cases?
  Low dimension, computations and some theory show small gap.
  Recent negative results in very high dimension, though (Blekherman)
- Isn't it a very special formulation?
  No, we can approximate *any* semialgebraic problem!
- How? And how do you solve them?
  OK, I'll tell you. But first, an example!

SOS programs: questions

- Why not just use nonnegative polynomials?
  While convex, unfortunately it's NP-hard ;(
- And is SOS any better?
  Yes, we can solve SOS programs in polynomial time
- Aren't we losing too much then?
  In several important cases (quadratic, univariate, etc), nonnegativity and SOS is the same thing.
- And in the other cases?
  Low dimension, computations and some theory show small gap.
  Recent negative results in very high dimension, though (Blekherman)
- Isn't it a very special formulation?
  No, we can approximate *any* semialgebraic problem!
- How? And how do you solve them?

OK, I'll tell you. But first, an example!

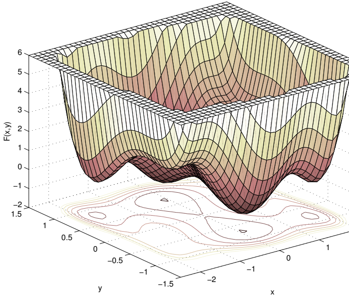
Global optimization

Given a multivariate polynomial, can we find the global minimum? Not convex. Many local minima. NP-hard. How to find good lower bounds?

- Find the largest γ s.t.

$$F(x, y) - \gamma \text{ is SOS.}$$

- If exact, can recover optimal solution.
- Surprisingly effective.



Often, the optimal γ is the true minimum.
Extensions to constrained case via representation theorems (Putinar/Lasserre) or the Positivstellensatz, yield *hierarchies* of relaxations.

SOS constraints are SDPs

"Gram matrix" method: $F(x)$ is SOS iff $F(x) = w(x)^T Q w(x)$, where $w(x)$ is a vector of monomials, and $Q \succeq 0$.
Let $F(x) = \sum f_\alpha x^\alpha$. Index rows and columns of Q by monomials. Then,

$$F(x) = w(x)^T Q w(x) \qquad \Leftrightarrow \qquad f_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta\gamma}$$

Thus, we have the SDP feasibility problem

$$f_\alpha = \sum_{\beta+\gamma=\alpha} Q_{\beta\gamma}, \qquad Q \succeq 0$$

SOS Example

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

An SDP with equality constraints. 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 $F(x,y) = \frac{1}{2}(2x^2 - 3y^2 + xy)^2 + \frac{1}{2}(y^2 + 3xy)^2$

Example 1: range of nonnegativity

For what range of values of a is the polynomial

$$P(x,y) = x^4 + y^2 - 4axy + (2a-3)x^2 + ax + 20$$

a sum of squares? Nonnegative?

For SOS, essentially:

```
sosprogram([x,y]); sosdecvar([a]);
P = x^4 + y^2 - 4*a*x*y + (2*a-3)*x^2 + a*x + 20 ;
sosineq(P); sossetobj(a);
sossolve; sosgetsol(a);
```

The solution: $a \in [-.94823, 1.42413]$ (numerically correct to 8+ digits).

Both are roots of the irreducible polynomial

$1613120 + 545280a - 1234772a^2 + 517544a^3 - 364251a^4 - 410208a^5 + 369408a^6 - 164224a^7 + 82176a^8$.

Interestingly, for SOS the range is the same as that for $P(x,y) \geq 0$.

Example 2: extremal polynomials

Given $n \geq 1$, define $p(x) = \sum_{i=0}^{n} a_i x^i$, and consider the problem:

$$\max_{a_0,\ldots,a_n} a_n \qquad \text{s.t.} \quad |p(x)| < 1 \; \forall x \in [-1, 1],$$

In words:

How large can the leading coefficient of a univariate polynomial be if the polynomial is unit-bounded in the $[-1, 1]$ interval?

What is the optimal value? What does the extreme $p(x)$ look like?

A QE problem with two blocks of quantifiers, in $n + 2$ variables:

$$(\exists a_0)(\exists a_1)\cdots(\exists a_{n-1})(\forall x)[x \geq -1 \wedge x \leq 1] \Rightarrow [a_n x_n + \cdots + a_0 \leq 1 \wedge a_n x_n + \cdots + a_0 \geq -1]$$

Can we solve it? For what values of n?

Example 2: extreme polynomials

Given $n \geq 1$, define $p(x) = \sum_{i=0}^{n} a_i x^i$, and consider the problem:

$$\max_{a_0,\ldots,a_n} a_n \qquad \text{s.t.} \quad |p(x)| < 1 \ \forall x \in [-1, 1],$$

In words:

How large can the leading coefficient of a univariate polynomial be if the polynomial is unit-bounded in the $[-1, 1]$ interval?

What is the optimal value? What does the extreme $p(x)$ look like?

A QE problem with two blocks of quantifiers, in $n + 2$ variables:

$$(\exists a_0)(\exists a_1)\cdots(\exists a_{n-1})(\forall x)[x \geq -1 \wedge x \leq 1] \Rightarrow [a_n x_n + \cdots + a_0 \leq 1 \wedge a_n x_n + \cdots + a_0 \geq -1]$$

Can we solve it? For what values of n?

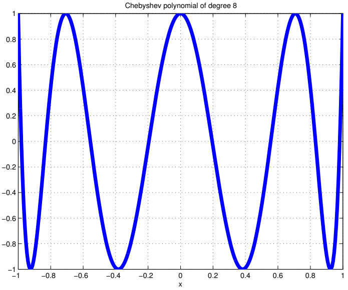
Example 2 (continued)

- By the Markov-Lukacs theorem,

$$p(x) \geq 0, \quad \forall\, x \in [-1,1] \qquad \Longleftrightarrow \qquad p(x) = s(x) + t(x)(1-x^2),$$

  where $s(x)$ and $t(x)$ are SOS.
- The optimal value is $a_n = 2^{n-1}$, and the optimal $p(x)$ are Chebyshev polynomials.

- Highly degenerate solution, $p(x)$ has many global minima.
- For $n \leq 12$, we solve it in ≈ 1 sec.
- For larger n, numerical issues become important.



Example: Joint spectral radius

Given a set of $n \times n$ matrices $\Sigma := \{A_1, \ldots, A_m\}$, what is the maximum "growth rate" that can be achieved by arbitrary switching?

$$\rho(\Sigma) := \limsup_{k \to +\infty} \max_{\sigma \in \{1,\ldots,m\}^k} ||A_{\sigma_k} \cdots A_{\sigma_2} A_{\sigma_1}||^{1/k}$$

Appeared in several different contexts: linear algebra (Rota-Strang 1960), wavelets (Daubechies-Lagarias 1992), switched linear systems, etc.

If $m = 1$, then $\rho(\{A_1\})$ is the *spectral radius* $\max |\lambda(A)|$.
If $m \geq 2$, determining if $\rho(\Sigma) \leq 1$ is undecidable (Blondel-Tsitsiklis 2000).

Upper bounds via polynomials

Thm: Let $p(x)$ be a strictly positive homogeneous multivariate polynomial of degree $2d$, that satisfies

$$\gamma^{2d}\, p(x) \,-\, p(A_i x) \geq 0 \qquad \forall x \in \mathbb{R}^n \quad i = 1, \ldots, m$$

Then, $\rho(\Sigma) \leq \gamma$.

A natural relaxation is obtained by replacing nonnegativity by SOS. Then:

Thm: The SOS relaxation satisfies:

$$\binom{n+d-1}{d}^{-\frac{1}{2d}} \rho_{SOS,2d} \leq \rho(\Sigma) \leq \rho_{SOS,2d}. \qquad (1)$$

Approximation ratio is *independent* of the number of matrices. As $d \to \infty$, the factor converges to 1.

Generalizations to constrained switching (Ahmadi et al. 2012).

Upper bounds via polynomials

Thm: Let $p(x)$ be a strictly positive homogeneous multivariate polynomial of degree $2d$, that satisfies

$$\gamma^{2d}\, p(x) - p(A_i x) \geq 0 \qquad \forall x \in \mathbb{R}^n \quad i = 1, \ldots, m$$

Then, $\rho(\Sigma) \leq \gamma$.
A natural relaxation is obtained by replacing nonnegativity by SOS. Then:
Thm: The SOS relaxation satisfies:

$$\binom{n+d-1}{d}^{-\frac{1}{2d}} \rho_{SOS,2d} \leq \rho(\Sigma) \leq \rho_{SOS,2d}. \tag{1}$$

Approximation ratio is *independent* of the number of matrices. As $d \to \infty$, the factor converges to 1.

Generalizations to constrained switching (Ahmadi et al. 2012).

SOS and SDP

Strong relationship between SOS programs and SDP.
In full generality, they are equivalent to each other.

- Semidefinite matrices are SOS quadratic forms.
- Conversely, can embed SOS polynomials into PSD cone.

However, they are a *very special* kind of SDP, with very rich algebraic and combinatorial properties.
Exploiting this structure is *crucial* in applications.
Both algebraic and numerical methods are required.

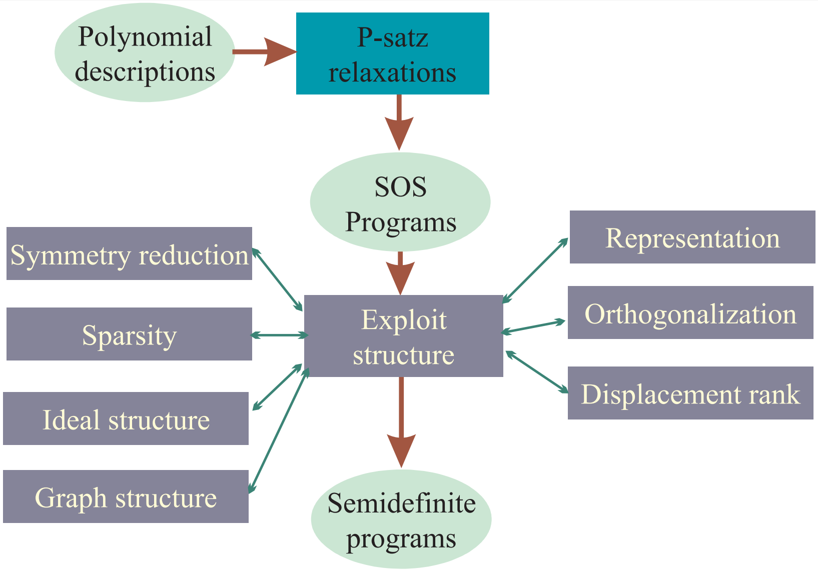
SOS and SDP

Strong relationship between SOS programs and SDP.
In full generality, they are equivalent to each other.

- Semidefinite matrices are SOS quadratic forms.
- Conversely, can embed SOS polynomials into PSD cone.

However, they are a *very special* kind of SDP, with very rich algebraic and combinatorial properties.
Exploiting this structure is *crucial* in applications.
Both algebraic and numerical methods are required.

Exploiting structure: algebraic and numerical



Perspectives, challenges, open questions

- Theory:
  - Better understanding of interaction between algebra and convexity ("convex algebraic geometry")
  - Minimum rank decompositions? Low-rank approaches?
  - Proof complexity, lower bounds, etc.
  - Connections with theoretical computer science
- Computation and numerical efficiency:
  - Specialized algorithms, better than SDP
  - Alternatives to interior point methods?
  - Increase numerical stability (better bases, splines, etc)
  - Representation issues: straight-line programs?
- Many more applications. . .

Summary

- A very rich class of optimization problems
- Methods have enabled many new applications
- Mathematical structure must be exploited for reliability and efficiency
- Combination of numerical and algebraic techniques.
- Fully algorithmic implementations

If you want to know more:

- Papers, slides, lecture notes, software, etc. at website: `www.mit.edu/~parrilo`

Thanks for your attention!

Summary

- A very rich class of optimization problems
- Methods have enabled many new applications
- Mathematical structure must be exploited for reliability and efficiency
- Combination of numerical and algebraic techniques.
- Fully algorithmic implementations

If you want to know more:

- Papers, slides, lecture notes, software, etc. at website:
  `www.mit.edu/~parrilo`

Thanks for your attention!