LP and SOCP-based Approximations to SOS

(initially funded by AFOSR YIP)

Amir Ali Ahmadi
Princeton, ORFE
Affiliated member of PACM, COS, MAE, CSML
Optimization over nonnegative polynomials

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

- **Optimization**
  - Lower bounds on polynomial optimization problems

  \[
  \min_x p(x) \\
  \text{s.t. } g_i(x) \geq 0
  \]

- **Statistics/ML**
  - Fitting a polynomial to data subject to shape constraints (e.g., convexity, or monotonicity)

  \[
  \frac{\partial p(x)}{\partial x_j} \geq 0, \forall x \in B
  \]

- **Control**
  - Stabilizing controllers

  \[
  \dot{x} = f(x)
  \]

  \[
  V(x) > 0, \\
  V(x) \leq \beta \Rightarrow \nabla V(x)^T f(x) < 0
  \]

  Implies that \( \{x \mid V(x) \leq \beta\} \) is in the region of attraction

[AAA, Curmei, Hall]
How to prove nonnegativity?

\[ p(x) = x_1^4 - 6x_1^3x_2 + 2x_1^3x_3 + 6x_1^2x_2^2 + 9x_1^2x_2 - 6x_1x_2x_3 - 14x_1x_2x_3^2 + 4x_1x_3^3 + 5x_3^4 - 7x_2^2x_3^2 + 16x_2^4 \]

\[ p(x) = (x_1^2 - 3x_1x_2 + x_1x_3 + 2x_3^2)^2 + (x_1x_3 - x_2x_3)^2 + (4x_2^2 - x_3^2)^2. \]

• Optimization over sum of squares (SOS) polynomials is a semidefinite program (SDP)!

• SDPs can be solved in poly-time to arbitrary accuracy.
Practical limitations of SOS

- **Scalability** is a nontrivial challenge!

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

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

where $z = [1, x_1, ..., x_n, x_1 x_2, ..., x_n^d]^T$ is the vector of monomials of degree up to $d$.

- 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 $\sim 1.2M$ decision variables and $\sim 200k$ constraints
Simple idea...

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

Not any set inside SOS would work!

1) sums of 4\textsuperscript{th} powers of polynomials
2) sums of 3 squares of polynomials

Both sets are clearly inside the SOS cone, but linear optimization over them is intractable.
dsos and sdsos polynomials (1/3)

**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 some nonnegative constants $\alpha_i, \beta_{ij}^+, \beta_{ij}^-$. 

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

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

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

---

**Note:** $DSOS_{n,d} \subseteq SDSOS_{n,d} \subseteq SOS_{n,d} \subseteq POS_{n,d}$
<table>
<thead>
<tr>
<th>Sum of squares (sos)</th>
<th>$p(x) = z(x)^T Q z(x), Q \succeq 0$</th>
<th>SDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSD cone: ${Q \mid Q \succeq 0}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DD cone: ${Q \mid Q_{ii} \geq \sum_{j \neq i}</td>
<td>Q_{ij}</td>
<td>, \forall i}$</td>
</tr>
<tr>
<td>SDD cone: ${Q \mid \exists$ diagonal $D$ with $D_{ii} &gt; 0$ s.t. $D Q D$ $dd}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Diagonally dominant sum of squares (dsos) | $p(x) = z(x)^T Q z(x), Q$ diagonally dominant (dd) | LP |

| Scaled diagonally dominant sum of squares (sdsos) | $p(x) = z(x)^T Q z(x), Q$ scaled diagonally dominant (sdd) | SOCP |
dsos and sdsos polynomials (3/3)

How to do better?
Method #1: \( r\text{-dsos} \) and \( r\text{-sdsos} \) polynomials (1/2)

**Defn.**

- A polynomial \( p \) is **\( r\text{-dsos} \)** if \( p(x) \cdot \left( \sum_i x_i^2 \right)^r \) is dsos.
- A polynomial \( p \) is **\( r\text{-sdsos} \)** if \( p(x) \cdot \left( \sum_i x_i^2 \right)^r \) is sdsos.

\[ p(x_1, x_2) = x_1^4 + x_2^4 + ax_1^3 x_2 + (1 - \frac{1}{2}a - \frac{1}{2}b)x_1^2 x_2^2 + 2bx_1 x_2^3 \]

(a) The LP-based \( r\text{-dsos} \) hierarchy.  
(b) The SOCP-based \( r\text{-sdsos} \) hierarchy.
Method #1: r-dsos and r-sdsos polynomials (2/2)

- r-dsos can outperform sos!

\[ p(x) = x_1^4 x_2^2 + x_2^4 x_3^2 + x_3^4 x_1^2 - 3 x_1^2 x_2^2 x_3^2 \]

is 1-dsos but not sos.

**Theorem:** Any even positive definite form is r-dsos for some r.

- Even forms include *copositive programming (and all problems in NP).*

**Theorem:** Any form can be made even, while preserving positivity, by doubling the number of variables and degree.

- Leads to arbitrarily tight lower bounds on any polynomial optimization problem (with a compact feasible set).

[AAA, Hall, *Math of OR*] (2018 INFORMS Young Researchers Prize)
Method #2: dsos/sdsos + change of basis (1/2)

\[ p(x) = x_1^4 - 6x_1^3x_2 + 2x_1^3x_3 + 6x_1^2x_2^2 + 9x_1^2x_2^2 - 6x_1^2x_2x_3 - 14x_1x_2x_3^2 + 4x_1x_3^3 + 5x_3^4 - 7x_2^2x_3^2 + 16x_2^4 \]

\[ p(x) = z^T(x)Qz(x) \]

\[ Q = \begin{pmatrix} 1 & -3 & 0 & 1 & 0 & 2 \\ -3 & 9 & 0 & -3 & 0 & -6 \\ 0 & 0 & 16 & 0 & 0 & -4 \\ 1 & -3 & 0 & 2 & -1 & 2 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 2 & -6 & 4 & 2 & 0 & 5 \end{pmatrix} \]

\[ z(x) = (x_1^2, x_1x_2, x_2^2, x_1x_3, x_2x_3, x_3^2)^T \]

\[ p(x) = \tilde{z}^T(x) \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} \tilde{z}(x) \]

\[ \tilde{z}(x) = \begin{pmatrix} 2x_1^2 - 6x_1x_2 + 2x_1x_3 + 2x_3^2 \\ x_1x_3 - x_2x_3 \\ x_2^2 - \frac{1}{4}x_3^2 \end{pmatrix} \]

Goal: iteratively improve \( z(x) \)

[AAA, Hall, *Contemporary Mathematics*]
Method #2: dsos/sdsos + change of basis (2/2)

\[ \begin{align*}
\text{LP} & \quad \begin{bmatrix} \max \ l(\tilde{P}) \\ \tilde{P}, Q \end{bmatrix} \\
& \text{st. } P(x) = z^T(x)Qz(x) \forall x \\
& Q \text{ dd} \\
\rightarrow \text{Optimal soln. } C^*_Q \\
\rightarrow \text{Cholesky: } C^*_Q = U^T U \\
\text{LP}_+ & \quad \begin{bmatrix} \max \ l(\tilde{P}) \\ \tilde{P}, Q \end{bmatrix} \\
& \text{st. } P(x) = z^T(x)U^TQUz(x) \forall x \\
& Q \text{ dd} \\
\end{align*} \]

\[ p_{a,b}(x_1, x_2) = 2x_1^4 + 2x_2^4 + ax_1^3x_2 + (1 - a)x_1^2x_2^2 + bx_1x_2^3 \]
Applications in control

(see paper for applications in statistics, finance, combinatorial and polynomial optimization)

Reminder

\[ \dot{x} = f(x, u) \]

Stability of equilibrium points

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

implies \( \{ x \mid V(x) \leq \beta \} \) is in the region of attraction (ROA)
Stabilizing the inverted $N$-link pendulum (2N states)

Runtime:

<table>
<thead>
<tr>
<th>$2N$ (# states)</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSOS</td>
<td>&lt; 1</td>
<td>0.72</td>
<td>6.72</td>
<td>7.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDSOS</td>
<td>&lt; 1</td>
<td>3.97</td>
<td>156.9</td>
<td>1697.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOS (SeDuMi)</td>
<td>&lt; 1</td>
<td>0.84</td>
<td>16.2</td>
<td>149.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOS (MOSEK)</td>
<td>&lt; 1</td>
<td></td>
<td></td>
<td></td>
<td>23676.5</td>
<td>∞</td>
<td>∞</td>
<td>∞</td>
<td>∞</td>
<td>∞</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1526.5</td>
<td>∞</td>
<td>∞</td>
<td>∞</td>
<td>∞</td>
<td>∞</td>
</tr>
</tbody>
</table>

ROA volume ratio:

<table>
<thead>
<tr>
<th>$2N$ (states)</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{\text{dsos}}/\rho_{\text{sos}}$</td>
<td>0.38</td>
<td>0.45</td>
<td>0.13</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>$\rho_{\text{adsos}}/\rho_{\text{sos}}$</td>
<td>0.88</td>
<td>0.84</td>
<td>0.81</td>
<td>0.79</td>
<td>0.79</td>
</tr>
</tbody>
</table>

[Majumdar, AAA, Tedrake, CDC]
Stabilizing ATLAS

- 30 states
- 14 control inputs
- Cubic dynamics

Done by SDSOS Optimization

[Majumdar, AAA, Tedrake, CDC]

https://github.com/spot-toolbox/spotless
Real-time barrier certificates

Collision avoidance

\[
\begin{align*}
\dot{x} &= f(x) \\
(f: \mathbb{R}^n \rightarrow \mathbb{R}^n)
\end{align*}
\]

\[SV: \text{needs safety verification} \]

\[U: \text{unsafe (or forbidden) set} \]

\[V(x) < 0, \forall x \in S \]

\[V(x) > 0, \forall x \in U \]

\[\dot{V}(x) = \langle \nabla V(x), f(x) \rangle \leq 0 \]

SDOS run time: \textasciitilde ms

[AAA, Majumdar, Optimization Letters]
Take-away message

There is plenty of room inside the SOS cone!

- Trade-offs between the different approximations?
- What algorithm to use when? Can ML tell us?
- How to exploit sparsity, symmetry, etc. on top of this?
- Real-time implementations?

But we have a long way to go!

Want to know more? aaa.princeton.edu
Backup slides...
What can DSOS/SDSOS do in theory?

\[ p(x) > 0, \forall x \in \{x \in \mathbb{R}^n \mid g_i(x) \geq 0, i = 1, \ldots, m\} \]

- Is there always an SOS proof?
  
  Yes, e.g. based on Putinar’s Psatz. (under a compactness assumption)

- Is there always an SDSOS proof?

- Is there always an DSOS proof?
  
  Yes! In fact, a much stronger statement is true.

If \( p(x) > 0, \forall x \in S, \) then \( p(x) = \sigma_0(x) + \sum_i \sigma_i(x)g_i(x), \) where \( \sigma_0, \sigma_i \) are sos
An optimization-free Positivstellensatz (1/2)

\[ p(x) > 0, \forall x \in \{ x \in \mathbb{R}^n | g_i(x) \geq 0, i = 1, \ldots, m \} \]

\[ 2d = \text{maximum degree of } p, g_i \]

\[ \iff \exists r \in \mathbb{N} \text{ such that} \]

\[ \left( f(v^2 - w^2) - \frac{1}{r} \left( \sum_i (v_i^2 - w_i^2)^2 \right)^d + \frac{1}{2r} \left( \sum_i (v_i^4 + w_i^4)^d \right) \right) \cdot (\sum_i v_i^2 + \sum_i w_i^2)^{r^2} \]

is \textbf{nonnegative coefficients},

where \( f \) is a form in \( n + m + 3 \) variables and of degree \( 4d \), which can be explicitly written from \( p, g_i \) and \( R \).

[AAA, Hall, Math of OR] (2018 INFORMS Young Researchers Prize)
An optimization-free Positivstellensatz (2/2)

\[ p(x) > 0 \text{ on } \{ x \mid g_i(x) \geq 0 \} \iff \exists r \in \mathbb{N} \text{ s.t. } \left( f(v^2 - w^2) - \frac{1}{r} \left( \sum_i (v_i^2 - w_i^2)^2 \right)^d + \frac{1}{2r} \left( \sum_i (v_i^4 + w_i^4) \right)^d \right) \cdot (\sum_i v_i^2 + \sum_i w_i^2)^r \]

has \( \geq 0 \) coefficients

- \( p(x) > 0 \) on \( \{ x \mid g_i(x) \geq 0 \} \iff f \) is pd

- **Result by Polya (1928):**
  
  \( f \) even and pd \( \Rightarrow \) \( \exists r \in \mathbb{N} \) such that \( f(z) \cdot (\sum_i z_i^2)^r \) has nonnegative coefficients.

- Make \( f(z) \) even by considering \( f(v^2 - w^2) \). We lose positive definiteness of \( f \) with this transformation.

- Add the positive definite term \( \frac{1}{2r} \left( \sum_i (v_i^4 + w_i^4) \right)^d \) to \( f(v^2 - w^2) \) to make it positive definite. **Apply Polya’s result.**

- The term \(- \frac{1}{r} \left( \sum_i (v_i^2 - w_i^2)^2 \right)^d \) ensures that the converse holds as well.

As a corollary, gives LP/SOCP-based converging hierarchies...

(Even forms with nonnegative coefficients are trivially dsos.)