Limitations on representing SOS cones with bounded size PSD blocks

James Saunderson

Electrical and Computer Systems Engineering, Monash University, Australia

December 12, 2017
Sums of squares

\[ p(x) = \sum_i [p_i(x)]^2 \implies p(x) \geq 0 \text{ for all } x \]

- sufficient condition for global nonnegativity
- generic tool for constructing convex optimization formulations/relaxations
- **Key observation:** \( \text{SOS}_{n,2d} \) has semidefinite description
Sums of squares

\[ p(x) = \sum_i [p_i(x)]^2 \implies p(x) \geq 0 \text{ for all } x \]

- **tractable?** sufficient condition for global nonnegativity
- **generic tool** for constructing convex optimization formulations/relaxations
- **Key observation:** \( \text{SOS}_{n,2d} \) has semidefinite description
Scalability: use SDPs with only small blocks

Inner approximations to SOS cone:

- DSOS: linear programming formulation (1 \times 1 blocks)
- SDSOS: second-order cone formulation (2 \times 2 blocks)

\[ p \text{ SDSOS} \iff p(x) = v_d(x)^T G v_d(x) \]

where \( G \) is “scaled diagonally dominant”
Scalability: use SDPs with only small blocks

Inner approximations to SOS cone:
- **DSOS**: linear programming formulation (1 × 1 blocks)
- **SDSOS**: second-order cone formulation (2 × 2 blocks)

\[ p \text{ SDSOS} \iff p(x) = v_d(x)^T G v_d(x) \]

where \( G \) is “scaled diagonally dominant”

Equivalently: there exist 2 × 2 psd matrices \( G_{\{i,j\}} \) s.t.

\[
G = \sum_{i<j} E_{\{i,j\}} G_{\{i,j\}} E_{\{i,j\}}^T
\]

Solution time for SDPs with (small) bounded blocks more like LP than general SDP
Challenge: what can be done with small blocks?

DSOS and SDSOS:

▶ particular strategies for approximating SOS cones with sets that can be described using small SDP blocks

Can we find

▶ Better approximations with fewer small blocks?
▶ Exact formulations of SOS cones using only small blocks?

How to reason about all possible SDP formulations with small blocks?
Lifts of convex sets

**Definition:** A convex set $C$ has a $K$-lift if there is an affine subspace $L$ and linear map $\pi$ such that

$$C = \pi(K \cap L)$$

If $C$ has a $K$-lift then linear optimization problems over $C$ can be formulated as conic programs over $K$. 
Lifts with small blocks: $\left(S^2_+\right)^p$-lifts

**Cone:** product of $2 \times 2$ PSD cones

$$\left(S^2_+\right)^p := S^2_+ \times \cdots \times S^2_+ \ (p \text{ terms})$$

For a convex set:

$$\left(S^2_+\right)^p\text{-lift} \iff \text{LMI description with } 2 \times 2 \text{ blocks}$$

All basic ideas generalize to bounded block size case

**Examples:**

- $n \times n$ scaled diag. dominant matrices: has $\left(S^2_+\right)^{(n/2)}$-lift

- $\{X \in S^3_+ : X_{11} = X_{22}\}$ has $\left(S^2_+\right)^2$-lift
  (chordal sparsity after congruence transformation)
Some related work

Lifts using $1 \times 1$ blocks $\iff$ linear prog. descriptions

- **Existence easy:** $C$ has LP lift if and only if $C$ a polyhedron
- **Main effort:** lower bounds on size of lifts
  - Connection with nonnegative rank: Yannakakis (1991)
  - Correlation/CUT/TSP polytope: Fiorini et al. (2012)
  - Matching polytope: Rothvoß (2013)
Some related work

Lifts using $1 \times 1$ blocks $\leftrightarrow$ linear prog. descriptions

- **Existence easy:** $C$ has LP lift if and only if $C$ a polyhedron
- **Main effort:** lower bounds on size of lifts
  - Connection with nonnegative rank: Yannakakis (1991)
  - Correlation/CUT/TSP polytope: Fiorini et al. (2012)
  - Matching polytope: Rothvoß (2013)

No restriction on block size $\leftrightarrow$ general SDP descriptions

- Many constructions (including SOS cones)
- **Scheiderer (2017)**
  \[ \text{PSD}_{n,d} \text{ has } S^p_+-\text{lift if and only if } \text{PSD}_{n,d} = \text{SOS}_{n,d} \]
Some related work

Lifts using $1 \times 1$ blocks $\leftrightarrow$ linear prog. descriptions

- **Existence easy:** $C$ has LP lift if and only if $C$ a polyhedron
- **Main effort:** lower bounds on **size** of lifts
  - Connection with nonnegative rank: Yannakakis (1991)
  - Correlation/CUT/TSP polytope: Fiorini et al. (2012)
  - Matching polytope: Rothvoß (2013)

No restriction on block size $\leftrightarrow$ general SDP descriptions

- Many constructions (including SOS cones)
- **Scheiderer (2017)**
  
  $\text{PSD}_{n,d}$ has $\mathcal{S}^p_+$-lift if and only if $\text{PSD}_{n,d} = \text{SOS}_{n,d}$

Very little known about obstructions to representability with small blocks
Question: For which \((n, d)\) does \(\text{SOS}_{n,d}\) have an \((S^2_+)^p\)-lift?

- \((n, d) = (1, 2)\) (trivial)
- Are there any other cases with \((S^2_+)^p\)-lifts?
Fawzi’s result

Question: For which \((n, d)\) does \(\text{SOS}_{n,d}\) have an \((S_+^2)^p\)-lift?

- \((n, d) = (1, 2)\) (trivial)
- Are there any other cases with \((S_+^2)^p\)-lifts?

Fawzi (2016) The cone of non-negative univariate quartics does not have a \((S_+^2)^p\)-lift.

Corollaries: cannot describe using \(2 \times 2\) PSD blocks:
- \(\text{SOS}_{n,d}\) unless \((n, d) = (1, 2)\)
- \(n \times n\) PSD cone for \(n \geq 3\)
Associate **slack matrix** with convex cone $C$

$$S_{x,\ell} = \langle \ell, x \rangle$$

where

- $\ell$ linear functional non-negative on $C$
- $x$ an element of $C$

The slack matrix is entry-wise nonnegative.

**Lifts of** $C$ **correspond to structured factorizations of** $S$
Lifts of convex sets and $S^2_+$-rank

A nonnegative matrix $S$ has $S^2_+$-rank one if $\exists A_i, B_j \in S^2_+$ s.t.

$$S = \begin{bmatrix}
\langle A_1, B_1 \rangle & \langle A_1, B_2 \rangle & \cdots & \langle A_1, B_b \rangle \\
\langle A_2, B_1 \rangle & \langle A_2, B_2 \rangle & \cdots & \langle A_2, B_b \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle A_a, B_1 \rangle & \langle A_a, B_2 \rangle & \cdots & \langle A_a, B_b \rangle
d\end{bmatrix}$$

Definition: The $S^2_+$-rank of an entrywise nonnegative matrix $S$ is the smallest $p$ such that $S = S_1 + S_2 + \cdots + S_p$ where each $S_k$ has $S^2_+$-rank one.

Theorem [Gouveia, Parrilo, Thomas 2013]
If $C$ has a proper $(S^2_+)^p$-lift then (any submatrix of) its slack matrix has has $S^2_+$-rank at most $p$. 
Slack matrix of non-negative univariate quartics

Indexed by non-neg. polynomials $p \in \text{SOS}_{1,4}$ and points $t \in \mathbb{R}$:

$$S_{p,t} = p(t) \geq 0$$

If $\text{SOS}_{1,4}$ had $(S^2_+)^p$-lift then for any non-negative quartics $p_1, \ldots, p_a$ and any points $t_1, \ldots, t_b \in \mathbb{R}$, could write

$$
\begin{bmatrix}
    p_1(t_1) & p_1(t_2) & \cdots & p_1(t_1) \\
    p_2(t_1) & p_2(t_2) & \cdots & p_2(t_2) \\
    \vdots & \vdots & \ddots & \vdots \\
    p_a(t_1) & p_a(t_2) & \cdots & p_a(t_b)
\end{bmatrix}
= S_1 + S_2 + \cdots + S_p
$$

where each $S_i$ has $S^2_+$-rank one
Define sequence of submatrices

For positive integers $1 \leq i_1 < i_2$ define

$$p_{\{i_1, i_2\}}(t) = [(i_1 - i_2)(i_1 - t)(i_2 - t)]^2$$

Define $\binom{k}{2} \times k$ submatrices of $S$ by

$$S_{\{i_1, i_2\}, j}^{(k)} = p_{\{i_1, i_2\}}(j)$$

for $1 \leq i_1 < i_2 \leq k$ and $1 \leq j \leq k$

Example:

$$S^{(3)} = \begin{pmatrix}
p_{\{1,2\}} & 1 & 2 & 3 \\
p_{\{1,3\}} & 0 & 0 & 4 \\
p_{\{2,3\}} & 0 & 4 & 0 \\
4 & 0 & 0 & 0
\end{pmatrix}$$
Show that $S^2_+-\text{rank}$ of $S^{(k)}$ grows without bound

Key ingredients:

- if $k' < k$ then $S^{(k')}$ a submatrix of $S^{(k)}$
- if $S_{ij} = 0$ and $S = S_1 + \cdots + S_p$ with non-negative terms then $[S_k]_{ij} = 0$ for all $k$
- if $S^2_+-\text{rank}$ one matrix has two zeros in a non-zero row then the corresponding columns are scalings of each other
Approximations?

▶ How well can we approximate SOS cones with cones having SDP representations with few small blocks?

▶ Even for polyhedral approximations \((1 \times 1\) blocks) how do approximation quality and size of lift relate?

▶ Can we find quantitative lower bounds? What do obstructions look like?
Sum of squares optimization

**Useful:** in control, combinatorial optimization, analysis of games, quantum information, ... 

**Challenge:** Natural SDP formulation scales poorly with increasing degree/number of variables

**Possibilities:**
- Algorithms that exploit structure (e.g., sparsity)
- Alternative certificates of non-negativity: DSOS, SDSOS can search for these via LP/SOCP
- Iterative methods based on DSOS and SDSOS
- Better approximations with small blocks(?)
Exploiting sparsity in first-order methods

SOS programs:
- Coefficient matching constraints very sparse
- Have additional ‘partial orthogonality’ structure
- Can solve and exploit this structure using ADMM-based first-order methods

CDCS: open-source MATLAB solver for partially decomposable conic programs (including SOS)
DSOS and SDSOS

Search over inner approximations to SOS cone:

- DSOS: diag. dominant Gram matrix (LP)
- SDSOD: scaled diag. dominant Gram matrix (SOCP)

Trade-off

- (S)DSOS inner approx. $\implies$ ‘weaker’ than regular SOS
- BUT can solve problems ‘higher’ in $r$-(S)DSOS hierarchy
Adaptive non-negativity certificates

Classical SOS:
- choose subspace(s) of functions to take sums of squares from (e.g., polynomials of degree at most $d$)
- Search for DSOS/SDSOS/SOS certificates

(S)DSOS column generation:
- Large dictionary of small subspaces of functions to take sums of squares from
- Each iteration, add useful subspace to the dictionary

(S)DSOS Cholesky change of basis:
- Each iteration, update subspace(s) of functions to take sums of squares from
- Don’t increase size of subspace, but improve it

Systematic study of such adaptive certificates?
References

Fawzi’s paper

CDC tutorial paper
‘Improving Efficiency and Scalability of Sum of Squares Optimization: Recent Advances and Limitations’ arxiv.org/abs/1710.01358

Thank you!