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

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

- sufficient condition for global nonnegativity
- generic tool for constructing convex optimization formulations/relaxations
- Key observation: $\mathrm{SOS}_{n,2d}$ has semidefinite description

Sums of squares

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

- tractable? sufficient condition for global nonnegativity
- generic tool for constructing convex optimization formulations/relaxations
- Key observation: $\mathrm{SOS}_{n,2d}$ has semidefinite description

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"

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

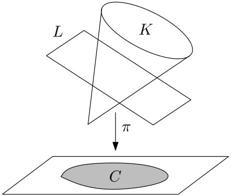
- Better approximations with fewer small blocks?
- Exact formulations of SOS cones using only small blocks?

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 π 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: $(\mathcal{S}_+^2)^p$-lifts

Cone: product of 2×2 PSD cones

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

For a convex set:

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

All basic ideas generalize to bounded block size case

Examples:

- $n \times n$ scaled diag. dominant matrices: has $(\mathcal{S}_+^2)^{\binom{n}{2}}$-lift
- $\{X \in \mathcal{S}_+^3 : X_{11} = X_{22}\}$ has $(\mathcal{S}_+^2)^2$-lift
  (chordal sparsity after congruence transformation)

Some related work

Lifts using 1×1 blocks $\longleftrightarrow$ 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 1×1 blocks $\longleftrightarrow$ 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 $\longleftrightarrow$ general SDP descriptions

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

Some related work

Lifts using 1×1 blocks $\longleftrightarrow$ 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 $\longleftrightarrow$ general SDP descriptions

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

Very little known about obstructions to representability with small blocks

Fawzi's result

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

- $(n, d) = (1, 2)$ (trivial)
- Are there any other cases with $(\mathcal{S}^2_+)^p$-lifts?

Fawzi's result

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

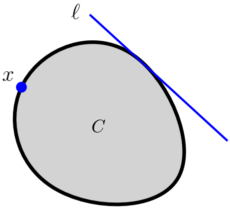
- $(n, d) = (1, 2)$ (trivial)
- Are there any other cases with $(\mathcal{S}_+^2)^p$-lifts?

Fawzi (2016) The cone of non-negative univariate quartics does not have a $(\mathcal{S}_+^2)^p$-lift.

Corollaries: cannot describe using 2×2 PSD blocks:

- $SOS_{n,d}$ unless $(n, d) = (1, 2)$
- $n \times n$ PSD cone for $n \geq 3$

Slack matrix



Associate slack matrix with convex cone C

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

where

- ℓ 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 $\mathcal{S}_+^2$-rank

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

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

Definition: The $\mathcal{S}_+^2$-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 $\mathcal{S}_+^2$-rank one.

Theorem [Gouveia, Parrilo, Thomas 2013]
If C has a proper $(\mathcal{S}_+^2)^p$-lift then (any submatrix of) its slack matrix has has $\mathcal{S}_+^2$-rank at most p.

Slack matrix of non-negative univariate quartics

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

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

If $SOS_{1,4}$ had $(\mathcal{S}^2_+)^p$-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 $\mathcal{S}^2_+$-rank one

Define sequence of submatrices

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

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

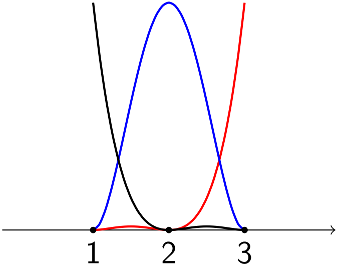
Define $\binom{k}{2} \times k$ submatrices of S by

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

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

Example:

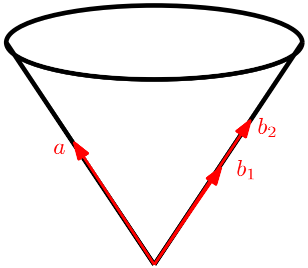
$$S^{(3)} = \begin{matrix} \\ p_{\{1,2\}} \\ p_{\{1,3\}} \\ p_{\{2,3\}} \end{matrix} \begin{matrix} 1 & 2 & 3 \\ \left[\begin{matrix} 0 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{matrix}\right] \end{matrix}$$



Show that $\mathcal{S}_+^2$-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 $\mathcal{S}_+^2$-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 (1×1 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(?)

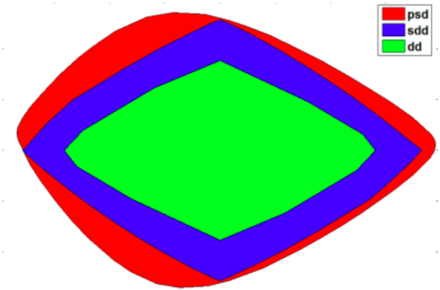
Exploiting sparsity in first-order methods

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

H. Fawzi, 'On representing the positive semidefinite cone using the second-order cone' arxiv.org/abs/1610.04901, 2016.

CDC tutorial paper

'Improving Efficiency and Scalability of Sum of Squares Optimization: Recent Advances and Limitations' arxiv.org/abs/1710.01358

Thank you!