Sparse sum-of-squares (SOS) optimization: A bridge between DSOS/SDSOS and SOS optimization for sparse polynomials

Yang Zheng[1], Giovanni Fantuzzi[2] and Antonis Papachristodoulou[3]

[1]School of Engineering and Applied Science, Harvard University

[2]Department of Aeronautics, Imperial College London

[3]Department of Engineering Science, University of Oxford

ACC 2019, Philadelphia, USA

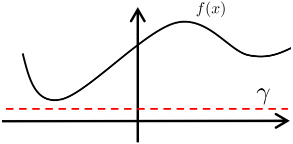
July 12, 2019

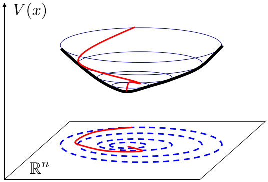
Outline

1. Introduction: sum-of-squares (SOS) optimization
2. Improving scalability via imposing structures on matrix Q
3. Bridging the gap via exploiting chordal sparsity
4. Numerical examples
5. Conclusion

Introduction: sum-of-squares optimization

$$\begin{array}{rl} \max & \gamma \\ \text{subject to} & f(x) - \gamma \geq 0, \forall x \end{array}$$



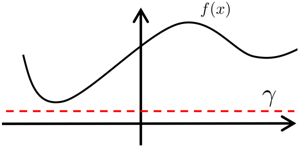


Optimization over nonnegative polynomials

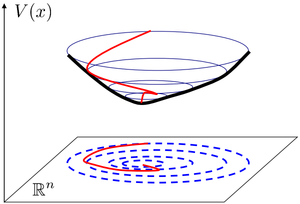
Motivation: Is $p(x) \geq 0$ over $\{x \in \mathbb{R}^n \mid g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$? This has a wide range of applications.

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

$$\begin{aligned} \max \quad & \gamma \\ \text{subject to} \quad & f(x) - \gamma \geq 0, \forall x \end{aligned}$$



- **Control:** Verifying asymptotic stability, finding region of attractions *etc.*



- Bernard, Lasserre Jean. Moments, positive polynomials and their applications. Vol. 1. World Scientific, 2009.
- Blekherman, Grigoriy, Pablo A. Parrilo, and Rekha R. Thomas, eds. Semidefinite optimization and convex algebraic geometry. Society for Industrial and Applied Mathematics, 2012.

How would you prove nonnegativity? SOS $\rightarrow$ SDP

Main question: How to verify a polynomial $p(x)$ is non-negative $\forall x \in \mathbb{R}^n$?

This question is Not easy! (In fact, NP-hard for degree ≥ 4)

- **Sum-of-squares (SOS) polynomials**: $p(x)$ can be represented as a sum of finite squared polynomials $f_i(x), i = 1, \ldots, m$

$$p(x) = \sum_{i=1}^{m} f_i^2(x),$$

- **SDP characterization (Parrilo, Lasserre etc.):** $p(x)$ is SOS if and only if there exists $Q \succeq 0$,

$$p(x) = v_d(x)^T Q v_d(x).$$

where $v_d(x) = [1, x_1, x_2, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_n^d]^T$ is the standard monomial basis.

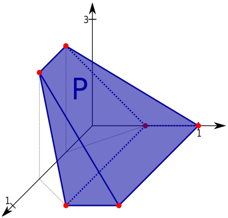
- **Scalability issue:** The size of the resulting SDP is

$$\binom{n+d}{d} \times \binom{n+d}{d},$$

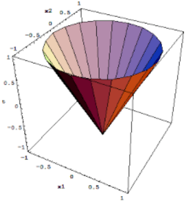
e.g. $n = 10, d = 4 \rightarrow 1001$.

Improving scalability via imposing structures on matrix Q

Linear program



second-order cone program



Imposing structures on matrix Q

SOS polynomials

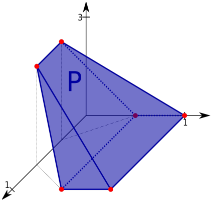
$$p(x) = v_d(x)^T Q v_d(x), \qquad Q \succeq 0$$

Inner approximations (Amadhi & Anirudha, 2019)

1. Q is diagonally dominant (dd).

$$Q_{ii} \geq \sum_{j=1, j\neq i}^{N} Q_{ij}, \forall i$$

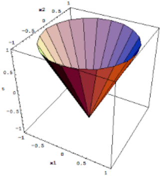
$\rightarrow$ linear program



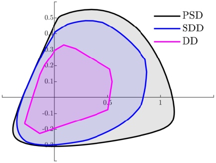
2. Q is scaled diagonally dominant (sdd).

$\exists$ diagonal $D \succeq 0$, s.t.DQD is dd $\rightarrow$

second-order cone program.



3. Other methods based on symmetry/sparsity of the polynomial $p(x)$, *e.g.*, Gatermann & Parrilo, 2004; Waki, Kim, Kojima, & Muramatsu, 2006.



- Ahmadi, Amir Ali, and Anirudha Majumdar. "DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization." SIAM Journal on Applied Algebra and Geometry 3.2 (2019): 193-230.

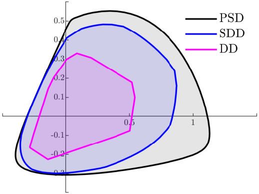
The gap between DSOS/SDSOS and SOS

A brief summary

- **SOS:** $p(x) = v_d(x)^T Q v_d(x)$: Q is PSD $\longrightarrow$ SDP
- **SDSOS:** $p(x) = v_d(x)^T Q v_d(x)$: Q is sdd $\longrightarrow$ SOCP
- **DSOS:** $p(x) = v_d(x)^T Q v_d(x)$: Q is dd $\longrightarrow$ LP

Another viewpoint

- **SDP**: involves PSD constraints of dimension $N \times N$
- **SOCP**: involves PSD constraints of dimension 2×2
- **LP**: involves PSD constraints of dimension 1×1

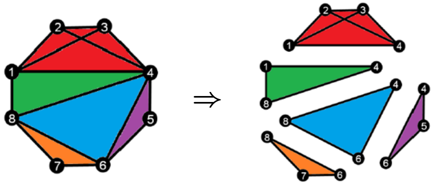


What is missing? How about problems that involve PSD constraints of dimension $k \times k$, where $1 \leq k \leq N$

- One approach: factor-width k matrices (Boman, et al. 2005) $\longrightarrow$ Not practical $\binom{n}{k} = \mathcal{O}(n^k)$
- **Chordal decomposition** by exploiting problem sparsity $\longrightarrow$ the main topic today.

Bridging the gap via exploiting chordal sparsity

— Chordal decomposition



- Vandenberghe, Lieven, and Martin S. Andersen. "Chordal graphs and semidefinite optimization." Foundations and Trends in Optimization 1.4 (2015): 241-433.

Sparsity in polynomials

- **Question:** How to describe the sparsity in a polynomial

$$p(x) = x_1^2 + x_1x_2 + x_2^2x_3^2$$

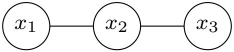
- **Correlative sparsity pattern:** (Waki et al, 2006) a symmetric matrix $\mathrm{csp}(p) \in \mathbb{S}^n$

$$[\mathrm{csp}(p)]_{ij} = \begin{cases} 1, & \text{if } i = j \text{ or } \exists \alpha \mid \alpha_i, \alpha_j \geq 1 \text{ and } c_\alpha \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

For example, we have

$$\mathrm{csp}(p(x) = x_1^2 + x_1x_2 + x_2^2x_3^2) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

- Describe the pattern of $\mathrm{csp}(p) \in \mathbb{S}^n$ as an undirected graph

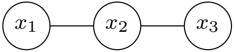


Define a set of sparse SOS polynomials as

$$\textit{SOS}_{n,2d}(\mathcal{E}) := \{p(x) \mid \operatorname{csp}(p) \in \mathbb{S}^n(\mathcal{E}, 0)\} \cap \textit{SOS}_{n,2d}.$$

Question: How to use the graph information?

$$p(x) = x_1^2 + x_1x_2 + x_2^2x_3^2$$



- Motivation from matrix decomposition (a special case of chordal decomposition)

$$\underbrace{\begin{bmatrix} * & * & 0 \\ * & * & * \\ 0 & * & * \end{bmatrix}}_{\succeq 0} = \underbrace{\begin{bmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\succeq 0} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{bmatrix}}_{\succeq 0}$$

- **Imposing a sparsity structure on** Q: A sparse polynomial $p(x) \in \textit{SOS}_{n,2d}(\mathcal{E})$ does not necessarily means a sparse Q

$$p(x) = \sum_{\beta,\gamma \in \mathbb{N}_d^n} Q_{\beta,\gamma} x^{\beta+\gamma} = \sum_{\alpha \in \mathbb{N}_{2d}^n} \left(\sum_{\beta+\gamma=\alpha} Q_{\beta,\gamma} \right) x^{\alpha}.$$

Decomposition in sparse polynomials

Recall

$$p(x) = v_d^{\mathsf{T}}(x) Q v_d(x) = \sum_{\beta,\gamma \in \mathbb{N}_d^n} Q_{\beta,\gamma} x^{\beta+\gamma}, \qquad Q \succeq 0.$$

Our key idea: Imposing sparsity in matrix Q:

- Define as subset of

$$\mathit{SSOS}_{n,2d}(\mathcal{E}) \subset \mathit{SOS}_{n,2d}(\mathcal{E})$$

  by imposing $Q_{\beta,\gamma} = 0$ if $x^{\beta+\gamma}$ violates the correlative sparsity pattern $\mathcal{E}$.

Example 1: Quadratic polynomials

$$p(x) = x_1^2 + 2x_1x_2 + 2x_2^2 + 2x_2x_3 + x_3^2 = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \underbrace{\begin{bmatrix} * & * & 0 \\ * & * & * \\ 0 & * & * \end{bmatrix}}_{Q \succeq 0} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \left(\underbrace{\begin{bmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\succeq 0} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{bmatrix}}_{\succeq 0} \right) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= (x_1 + x_2)^2 + (x_2 + x_3)^2 \in \mathit{SSOS}_{3,2}(\mathcal{E}).$$

Decomposition in sparse polynomials

Our key idea: Imposing sparsity in matrix Q:

- Define as subset of

$$\mathit{SSOS}_{n,2d}(\mathcal{E}) \subset \mathit{SOS}_{n,2d}(\mathcal{E})$$

by imposing $Q_{\beta,\gamma} = 0$ if $x^{\beta+\gamma}$ violates the correlative sparsity pattern $\mathcal{E}$.

Example 2: Quartic polynomials

$$p(x) = 1 + x_1^2 + 2x_1x_2 + x_2^2 + x_2^2x_3^2$$

$$= \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1^2 \\ x_1x_2 \\ x_1x_3 \\ x_2^2 \\ x_2x_3 \\ x_3^2 \end{bmatrix}^{\mathsf{T}} \underbrace{\begin{bmatrix} * & * & * & * & * & * & * & * & * & * \\ * & * & * & 0 & * & * & 0 & * & 0 & 0 \\ * & * & * & * & * & * & * & * & * & * \\ * & 0 & * & * & 0 & 0 & 0 & * & * & * \\ * & * & * & 0 & * & * & 0 & * & 0 & 0 \\ & & & & \cdots & & & & & \\ & & & & \cdots & & & & & \\ & & & & \cdots & & & & & \\ & & & & \cdots & & & & & \\ & & & & \cdots & & & & & \end{bmatrix}}_{Q \succeq 0} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1^2 \\ x_1x_2 \\ x_1x_3 \\ x_2^2 \\ x_2x_3 \\ x_3^2 \end{bmatrix}$$

$$= 1 + (x_1 + x_2)^2 + (x_2x_3)^2 \in \mathit{SSOS}_{3,4}(\mathcal{E}).$$

Decomposition in sparse polynomials

Result 1: Sparse polynomial decomposition

Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be a chordal graph with maximal cliques $\mathcal{C}_1, \ldots, \mathcal{C}_t$. Then,

$$p(x) \in \mathit{SSOS}_{n,2d}(\mathcal{E}) \Longleftrightarrow p(x) = \sum_{k=1}^{t} p_k(E_{\mathcal{C}_k} x),$$

where $p_k(E_{\mathcal{C}_k} x)$ is SOS that depends on a subset of the variable x.

- The proof is easy: since Q is sparse by definition, then applying chordal decomposition leads to the result;

$$\begin{aligned} p(x) &= v_d^{\mathsf{T}}(x) Q v_d(x) \\ &= v_d^{\mathsf{T}}(x)(Q_1 + Q_2 + \ldots + Q_t) v_d(x) \\ &= \sum_{k=1}^{t} v_d^{\mathsf{T}}(x) Q_k v_d(x) = \sum_{k=1}^{t} p_k(E_{\mathcal{C}_k} x). \end{aligned}$$

- This result is the same as the correlative sparsity technique by Waki et al. 2006.

Summary: LP/SOCP/SDP

Result 2: A hierarchy of inner approximations:

For any sparsity pattern $\mathcal{E}$, we have the following inclusion relationship

$$DSOS_{n,2d}(\mathcal{E}) \subset SDSOS_{n,2d}(\mathcal{E}) \subset SSOS_{n,2d}(\mathcal{E}) \subseteq SOS_{n,2d}(\mathcal{E})$$

- Proof idea: if a matrix is (scaled) diagonally dominant, then it is still (scaled) diagonally dominant when replacing any off-diagonal elements with zeros.

A brief summary (scalability):

$$DSOS_{n,2d}(\mathcal{E}) \longrightarrow \text{LP (PSD cones: } 1 \times 1)$$
$$SDSOS_{n,2d}(\mathcal{E}) \longrightarrow \text{SOCP (PSD cones: } 2 \times 2)$$
$$SSOS_{n,2d}(\mathcal{E}) \longrightarrow \text{SDP with smaller PSD cones of } k \times k$$
$$SOS_{n,2d}(\mathcal{E}) \longrightarrow \text{SDP with a PSD cone of } N \times N$$

Solution quality: $\mathcal{P}_{\text{dsos}}, \mathcal{P}_{\text{sdsos}}$ and $\mathcal{P}_{\text{ssos}}$ are a sequence of inner approximations with increasing accuracy to the SOS problem $\mathcal{P}_{\text{sos}}$, meaning that

$$f^*_{\text{dsos}} \geq f^*_{\text{sdsos}} \geq f^*_{\text{ssos}} \geq f^*_{\text{sos}},$$

Implementations and numerical comparison

Packages

- SOS optimization: SOSTOOLS, YALMIP
- DSOS/SDSOS optimization: SPOTLESS
- Chordal decomposition: YALMIP (we adapted the option of correlative sparsity technique)
- SDP solver: Mosek

Numerical examples and applications

- Polynomial optimization: eigenvalues bounds on polynomial matrices
- Control application: finding Lyapunov functions

Example 1: Polynomial optimization problems

Eigenvalue bounds on matrix polynomials

$$\begin{array}{ll} \min_{\gamma} & \gamma \\ \text{subject to} & P(x) + \gamma I \succeq 0, \forall x \in \mathbb{R}^2 \end{array}$$

where $n = 2, 2d = 2$, the polynomial is randomly generated. $P(x)$ has an arrow pattern.

Table: Optimal value γ

Dimension r	10	20	30	40	50	60	70	80
SOS	1.447	4.813	5.917	4.154	21.61	10.09	7.364	10.19
SSOS	1.454	4.878	5.917	4.498	21.64	12.71	7.558	11.39
SDSOS	40.1	279.3	1 254.4	145.5	762.8	1 521.1	1 217.3	598.0
DSOS	**	**	**	**	**	**	**	**

**: The program is infeasible.

Example 1: Polynomial optimization problems

Eigenvalue bounds on matrix polynomials

$$\begin{aligned} \min_{\gamma} \quad & \gamma \\ \text{subject to} \quad & P(x) + \gamma I \succeq 0, \forall x \in \mathbb{R}^2 \end{aligned}$$

where $n = 2, 2d = 2$, the polynomial is randomly generated. $P(x)$ has an arrow pattern.

Table: CPU time (in seconds) required by Mosek (*not very fair*)

Dimension r	10	20	30	40	50	60	70	80
SOS	0.30	1.33	6.64	27.3	108.1	308.7	541.3	1 018.6
SSOS	0.34	0.34	0.35	0.35	0.33	0.32	0.32	0.33
SDSOS	0.47	0.63	1.09	1.29	2.67	3.70	4.40	6.02
DSOS	**	**	**	**	**	**	**	**

**: The program is infeasible.

Example 2: Finding Lyapunov functions

Control application: finding Lyapunov functions

- Consider a dynamical system with a banded pattern

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2), & g_1(x) &= \gamma - x_1^2 \geq 0 \\ \dot{x}_2 &= f_2(x_1, x_2, x_3), & g_2(x) &= \gamma - x_2^2 \geq 0 \\ &\vdots \\ \dot{x}_n &= f_n(x_{n-1}, x_n), & g_2(x) &= \gamma - x_n^2 \geq 0 \end{aligned}$$

- Generate locally stable systems of degree three;
- Consider a polynomial Lyapunov function of degree two with a banded pattern

$$V(x) = V_1(x_1, x_2) + V_2(x_1, x_2, x_3) + \ldots + V_n(x_{n-1}, x_n)$$

- Then, we consider the following SOS program

$$\begin{aligned} \text{Find} \quad & V(x), r_i(x) \\ \text{subject to} \quad & V(x) - \epsilon(x^T x) \text{ is SOS} \\ & -\langle \nabla V(x), f(x) \rangle - \sum_{i=1}^{n} r_i(x) g_i(x) \text{ is SOS} \\ & r_i(x) \text{ is SOS}, i = 1, \ldots, n. \end{aligned}$$

Control application: finding Lyapunov functions

Table: CPU time (in seconds) required by Mosek (*not very fair*)

n	10	15	20	30	40	50
SOS	1.29	18.44	247.84	*	*	*
SSOS	0.55	0.68	0.71	0.83	1.04	1.17
SDSOS	0.71	1.76	4.47	32.21	85.99	257.20
DSOS	0.70	1.42	3.58	35.12	73.64	324.32

*: Out of memory.

Conclusion

Take-home message

- **Message 1: Imposing structures on matrix Q:**

$$p(x) = v_d(x)^T Q v_d(x), \qquad Q \succeq 0$$

Different choices lead to different inner approximations.

- **Message 2: A hierarchy of inner approximations**: bridging the gap

$$\mathit{DSOS}_{n,2d}(\mathcal{E}) \subset \mathit{SDSOS}_{n,2d}(\mathcal{E}) \subset \mathit{SSOS}_{n,2d}(\mathcal{E}) \subseteq \mathit{SOS}_{n,2d}(\mathcal{E})$$

Maintain the correlative sparsity pattern of $p(x)$ by carefully imposing a sparsity pattern on Q

$DSOS_{n,2d}(\mathcal{E})$	$\longrightarrow$	LP (PSD cones: 1×1)
$SDSOS_{n,2d}(\mathcal{E})$	$\longrightarrow$	SOCP (PSD cones: 2×2)
$SSOS_{n,2d}(\mathcal{E})$	$\longrightarrow$	SDP with smaller PSD cones of $k \times k$
$SOS_{n,2d}(\mathcal{E})$	$\longrightarrow$	SDP with a PSD cone of $N \times N$

Future work: Exploit the sparsity in the degree of polynomials; Maintain the sparsity structure in the applications of SOS optimization.

Thank you for your attention!

Q & A

- Zheng, Yang, Giovanni Fantuzzi, and Antonis Papachristodoulou. "Sparse sum-of-squares (SOS) optimization: A bridge between DSOS/SDSOS and SOS optimization for sparse polynomials." arXiv preprint arXiv:1807.05463 (2018).