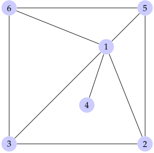
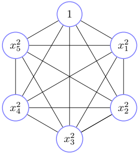
Sparse polynomial optimization

Victor Magron
LAAS CNRS

https://homepages.laas.fr/vmagron/courses.html

MINT Summer School "Moments, Positive Polynomials and their Applications" 28 June 2022





What is sparse polynomial optimization?

Looks like a regular polynomial optimization problem (POP):

$$\begin{aligned} \inf \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geqslant 0\} \end{aligned}$$

What is sparse polynomial optimization?

Looks like a regular polynomial optimization problem (POP):

$$\begin{aligned} \inf \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geqslant 0\} \end{aligned}$$

💡 But the input data f, g_j are "SPARSE"!

What is sparse polynomial optimization?

Looks like a regular polynomial optimization problem (POP):

$$\begin{aligned} \inf \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geqslant 0\} \end{aligned}$$

But the input data f, g_j are "SPARSE"!

Correlative sparsity: few products between each variable and the others in f, g_j

What is sparse polynomial optimization?

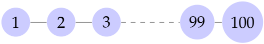
Looks like a regular polynomial optimization problem (POP):

$$\begin{aligned} \inf \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geqslant 0\} \end{aligned}$$

But the input data f, g_j are "SPARSE"!

Correlative sparsity: few products between each variable and the others in f, g_j

$\rightsquigarrow f(\mathbf{x}) = x_1 x_2 + x_2 x_3 + \ldots x_{99} x_{100}$



What is sparse polynomial optimization?

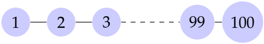
Looks like a regular polynomial optimization problem (POP):

$$\begin{aligned} \inf \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geqslant 0\} \end{aligned}$$

But the input data f, g_j are "SPARSE"!

Correlative sparsity: few products between each variable and the others in f, g_j

$\rightsquigarrow f(\mathbf{x}) = x_1x_2 + x_2x_3 + \dots x_{99}x_{100}$



Term sparsity: few terms in f, g_j

What is sparse polynomial optimization?

Looks like a regular polynomial optimization problem (POP):

$$\begin{aligned} \inf \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geqslant 0\} \end{aligned}$$

But the input data f, g_j are “SPARSE”!

Correlative sparsity: few products between each variable and the others in f, g_j

$\rightsquigarrow f(\mathbf{x}) = x_1 x_2 + x_2 x_3 + \ldots x_{99} x_{100}$

Term sparsity: few terms in f, g_j

$\rightsquigarrow f(\mathbf{x}) = x_1^{99} x_2 + x_1 x_2^{100}$

Where do we find sparse POPs?

Where do we find sparse POPs?

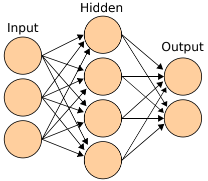
Everywhere (almost)!

Where do we find sparse POPs?

Everywhere (almost)!

Deep learning

$\rightsquigarrow$ robustness, computer vision

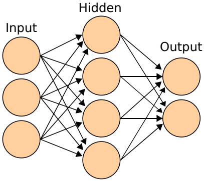


Where do we find sparse POPs?

Everywhere (almost)!

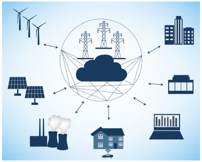
Deep learning

⇝ robustness, computer vision



Power systems

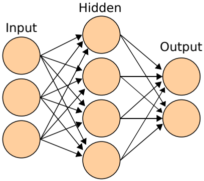
⇝ AC optimal power-flow, stability



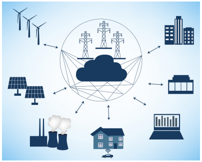
Where do we find sparse POPs?

Everywhere (almost)!

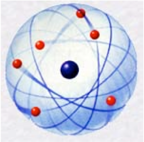
Deep learning
⇝ robustness, computer vision



Power systems
⇝ AC optimal power-flow, stability



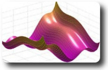
Quantum Systems
⇝ condensed matter, entanglement



The Moment-SOS Hierarchy for POP

NP-hard NON CONVEX Problem $f_{\min} = \inf f(\mathbf{x})$

Theory

(Primal)		(Dual)
$\inf \int f \, d\mu$		$\sup \lambda$
with μ proba $\Rightarrow$	INFINITE LP	$\Leftarrow$ with $f - \lambda \geqslant 0$

The Moment-SOS Hierarchy for POP

NP-hard NON CONVEX Problem $f_{\min} = \inf f(\mathbf{x})$

Practice

(Primal **Relaxation**)

moments $\int \mathbf{x}^\alpha \, d\mu$

finite number $\Rightarrow$

SDP

(Dual **Strengthening**)

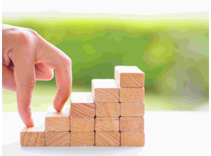
$f - \lambda =$ **sum of squares**

$\Leftarrow$ **fixed** degree

LASSERRE'S HIERARCHY of **CONVEX PROBLEMS** $\uparrow f^*$

[Lasserre '01]

degree d & n vars $\Longrightarrow \binom{n+2d}{n}$ **SDP** VARIABLES



The Moment-SOS Hierarchy for POP

NP-hard NON CONVEX Problem $f_{\min} = \inf f(\mathbf{x})$

Practice

(Primal **Relaxation**)

(Dual **Strengthening**)

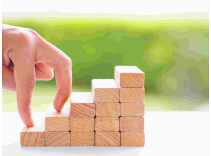
moments $\int \mathbf{x}^\alpha \, d\mu$

$f - \lambda =$ **sum of squares**

finite number $\Rightarrow$ **SDP** $\Leftarrow$ **fixed** degree

LASSERRE'S HIERARCHY of **CONVEX PROBLEMS** $\uparrow f^*$
[Lasserre '01]

degree d & n vars $\Longrightarrow \binom{n+2d}{n}$ **SDP** VARIABLES



HOW TO OVERCOME THIS NO-FREE LUNCH RULE?

The Moment-SOS Hierarchy for POP

NP-hard NON CONVEX Problem $f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$

- space $\mathcal{M}_+(\mathbf{X})$ of probability measures supported on $\mathbf{X}$
- quadratic module $\mathcal{Q}(\mathbf{X}) = \left\{ \sigma_0 + \sum_j \sigma_j g_j, \text{ with } \sigma_j \text{ SOS} \right\}$

Infinite-dimensional linear programs (LP)

(Primal) (Dual)

$$\begin{aligned} \inf \quad & \int_{\mathbf{X}} f \, d\mu \\ \text{s.t.} \quad & \mu \in \mathcal{M}_+(\mathbf{X}) \end{aligned} \quad = \quad \begin{aligned} \sup \quad & \lambda \\ \text{s.t.} \quad & \lambda \in \mathbb{R} \\ & f - \lambda \in \mathcal{Q}(\mathbf{X}) \end{aligned}$$

The Moment-SOS Hierarchy for POP

NP-hard NON CONVEX Problem $f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$

- Pseudo-moment sequences $\mathbf{y}$ up to order r
- Truncated quadratic module $\mathcal{Q}(\mathbf{X})_r$

Finite-dimensional semidefinite programs (SDP)

(Moment)		=	(SOS)	
inf	$\sum_\alpha f_\alpha y_\alpha$	=	sup	λ
s.t.	$\mathbf{M}_{r-r_j}(g_j \, \mathbf{y}) \succcurlyeq 0$		s.t.	$\lambda \in \mathbb{R}$
	$y_0 = 1$			$f - \lambda \in \mathcal{Q}(\mathbf{X})_r$

The Moment-SOS Hierarchy for POP

NP-hard NON CONVEX Problem $f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$

- Pseudo-moment sequences $\mathbf{y}$ up to order r
- Truncated quadratic module $\mathcal{Q}(\mathbf{X})_r$

Finite-dimensional semidefinite programs (SDP)

(Moment)		(SOS)
$\inf \sum_\alpha f_\alpha y_\alpha$	$=$	$\sup \lambda$
s.t. $\mathbf{M}_{r-r_j}(g_j \, \mathbf{y}) \succcurlyeq 0$		s.t. $\lambda \in \mathbb{R}$
$y_0 = 1$		$f - \lambda \in \mathcal{Q}(\mathbf{X})_r$

Primal-dual "SPARSE" variants?

Sparse SDP

Correlative sparsity in POP

Term sparsity in POP

Conclusion & further topics

Tutorial session

Sparse SDP

Correlative sparsity in POP

Term sparsity in POP

Conclusion & further topics

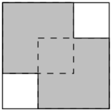
Tutorial session

Sparse matrices

Symmetric matrices indexed by graph vertices

Sparse matrices

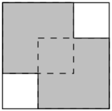
Symmetric matrices indexed by graph vertices



no edge between 1 and $3 \iff$ 0 entry in the $(1,3)$ entry

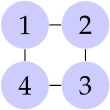
Sparse matrices

Symmetric matrices indexed by graph vertices



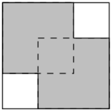
no edge between 1 and $3 \iff$ 0 entry in the $(1, 3)$ entry

cycle =



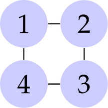
Sparse matrices

Symmetric matrices indexed by graph vertices



no edge between 1 and 3 $\Longleftrightarrow$ 0 entry in the $(1,3)$ entry

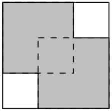
cycle =



chord = edge between two nonconsecutive vertices in a cycle

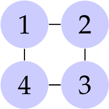
Sparse matrices

Symmetric matrices indexed by graph vertices



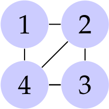
no edge between 1 and $3 \iff$ 0 entry in the $(1, 3)$ entry

cycle =



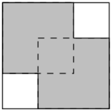
chord = edge between two nonconsecutive vertices in a cycle

chordal graph = all cycles of length $\geqslant$ 4 have at least one chord



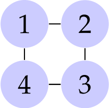
Sparse matrices

Symmetric matrices indexed by graph vertices



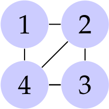
no edge between 1 and $3 \iff$ 0 entry in the $(1,3)$ entry

cycle =



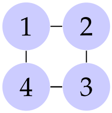
chord = edge between two nonconsecutive vertices in a cycle

chordal graph = all cycles of length $\geqslant$ 4 have at least one chord

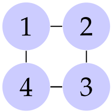


clique = a fully connected subset of vertices

Chordal extensions



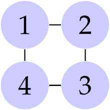
Chordal extensions



Fact

Any non-chordal graph can always be extended to a chordal graph, by adding appropriate edges

Chordal extensions

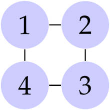


Fact

Any non-chordal graph can always be extended to a chordal graph, by adding appropriate edges

Chordal extension is not unique!

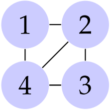
Chordal extensions

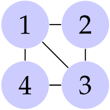


Fact

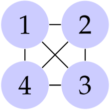
Any non-chordal graph can always be extended to a chordal graph, by adding appropriate edges

💡 Chordal extension is not unique!





approximately minimal



maximal

Theorem [Gavril '72, Vandenberghe & Andersen '15]

The maximal cliques of a chordal graph can be enumerated in linear time in the number of nodes and edges.

Running intersection property (RIP)

RIP Theorem for chordal graphs [Blair & Peyton '93]

For a chordal graph with maximal cliques $I_1, \ldots, I_p$:

$$\text{(RIP)} \quad \boxed{\forall k < p \quad I_{k+1} \cap \bigcup_{j \leqslant k} I_j \subseteq I_i \quad \text{for some } i \leqslant k}$$

(possibly after reordering)

Running intersection property (RIP)

RIP Theorem for chordal graphs [Blair & Peyton '93]

For a chordal graph with maximal cliques $I_1, \ldots, I_p$:

$$\text{(RIP)} \qquad \boxed{\forall k < p \quad I_{k+1} \cap \bigcup_{j \leqslant k} I_j \subseteq I_i \quad \text{for some } i \leqslant k}$$

(possibly after reordering)

💡 RIP always holds for $p = 2$

Running intersection property (RIP)

RIP Theorem for chordal graphs [Blair & Peyton '93]

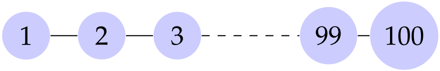
For a chordal graph with maximal cliques $I_1, \dots, I_p$:

$$\text{(RIP)} \qquad \boxed{\forall k < p \quad I_{k+1} \cap \bigcup_{j \leqslant k} I_j \subseteq I_i \quad \text{for some } i \leqslant k}$$

(possibly after reordering)

- RIP always holds for $p = 2$

- RIP holds for chains



Running intersection property (RIP)

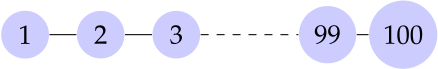
RIP Theorem for chordal graphs [Blair & Peyton '93]

For a chordal graph with maximal cliques $I_1, \ldots, I_p$:

$$\text{(RIP)} \quad \boxed{\forall k < p \quad I_{k+1} \cap \bigcup_{j \leqslant k} I_j \subseteq I_i \quad \text{for some } i \leqslant k}$$

(possibly after reordering)

- RIP always holds for $p = 2$
- RIP holds for chains

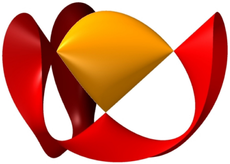


- RIP holds for numerous applications!

Semidefinite Programming (SDP)

$$\min_{\mathbf{y}} \quad \mathbf{c}^\top \mathbf{y}$$
$$\text{s.t.} \quad \sum_i \mathbf{F}_i \, y_i \succcurlyeq \mathbf{F}_0$$

- Linear cost $\mathbf{c}$
- Symmetric matrices $\mathbf{F}_0, \mathbf{F}_i$
- Linear matrix inequalities "$\mathbf{F} \succcurlyeq 0$" ($\mathbf{F}$ has nonnegative eigenvalues)



Spectrahedron

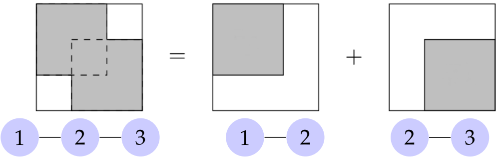
Sparse SDP matrices

Theorem [Griewank Toint '84, Agler et al. '88]

Chordal graph G with n vertices & maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries corresponding to edges of G

$\Longrightarrow \; Q_G = P_1{}^T Q_1 P_1 + P_2{}^T Q_2 P_2$ with $Q_k \succcurlyeq 0$ indexed by I_k



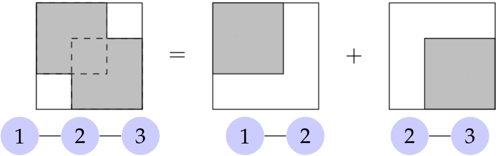
Sparse SDP matrices

Theorem [Griewank Toint '84, Agler et al. '88]

Chordal graph G with n vertices & maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries corresponding to edges of G

$\implies Q_G = {P_1}^T Q_1 P_1 + {P_2}^T Q_2 P_2$ with $Q_k \succcurlyeq 0$ indexed by I_k



What are P_1, P_2?

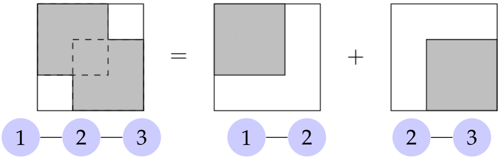
Sparse SDP matrices

Theorem [Griewank Toint '84, Agler et al. '88]

Chordal graph G with n vertices & maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries corresponding to edges of G

$\Longrightarrow \; Q_G = {P_1}^T Q_1 P_1 + {P_2}^T Q_2 P_2$ with $Q_k \succcurlyeq 0$ indexed by I_k



What are P_1, P_2? $P_1 \in \mathbb{R}^{|I_1| \times n}$

$$P(i,j) = \begin{cases} 1 & \text{if } I(i) = j \\ 0 & \text{otherwise} \end{cases}$$

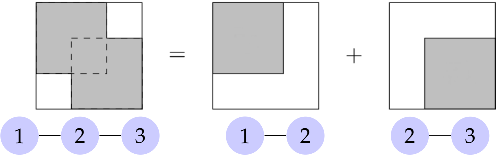
Sparse SDP matrices

Theorem [Griewank Toint '84, Agler et al. '88]

Chordal graph G with n vertices & maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries corresponding to edges of G

$\Longrightarrow \; Q_G = P_1{}^T Q_1 P_1 + P_2{}^T Q_2 P_2$ with $Q_k \succcurlyeq 0$ indexed by I_k



What are P_1, P_2? $P_1 \in \mathbb{R}^{|I_1| \times n}$

$$P(i,j) = \begin{cases} 1 & \text{if } I(i) = j \\ 0 & \text{otherwise} \end{cases}$$

$$I_1 = (1,2) \Rightarrow P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

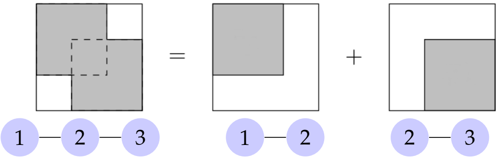
Sparse SDP matrices

Theorem [Griewank Toint '84, Agler et al. '88]

Chordal graph G with n vertices & maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries corresponding to edges of G

$\Longrightarrow \; Q_G = P_1{}^T Q_1 P_1 + P_2{}^T Q_2 P_2$ with $Q_k \succcurlyeq 0$ indexed by I_k



What are P_1, P_2? $P_1 \in \mathbb{R}^{|I_1| \times n}$

$$P(i,j) = \begin{cases} 1 & \text{if } I(i) = j \\ 0 & \text{otherwise} \end{cases}$$

$$I_1 = (1,2) \Rightarrow P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad I_2 = (2,3) \Rightarrow P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

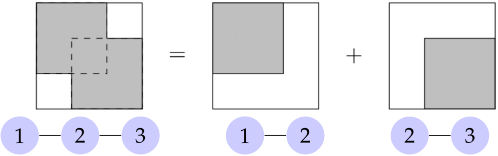
Sparse SDP matrices

Theorem [Griewank Toint '84, Agler et al. '88]

Chordal graph G with n vertices & maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries corresponding to edges of G

$\Longrightarrow \; Q_G = {P_1}^T Q_1 P_1 + {P_2}^T Q_2 P_2$ with $Q_k \succcurlyeq 0$ indexed by I_k



What are P_1, P_2? $P_1 \in \mathbb{R}^{|I_1| \times n}$

$$P(i,j) = \begin{cases} 1 & \text{if } I(i) = j \\ 0 & \text{otherwise} \end{cases}$$

$$I_1 = (1,2) \Rightarrow P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad I_2 = (2,3) \Rightarrow P_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

${P_1}^T Q_1 P_1$ inflates a $|I_1| \times |I_1|$ matrix Q_1 into a sparse $n \times n$ matrix

What is correlative sparsity?

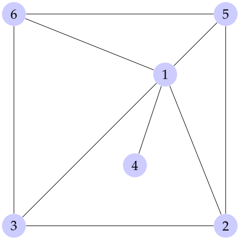
Exploit few links between **variables** [Lasserre, Waki et al. ’06]

$$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

Correlative sparsity pattern (csp) graph G

Vertices = $\{1, \ldots, n\}$

$(i, j) \in$ Edges $\Longleftrightarrow x_i x_j$ appears in f



What is correlative sparsity?

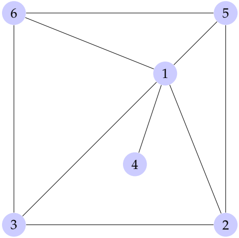
💡 Exploit few links between **variables** [Lasserre, Waki et al. ’06]

$$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

Correlative sparsity pattern (csp) graph G

Vertices = $\{1, \ldots, n\}$

$(i, j) \in$ Edges $\iff x_i x_j$ appears in f

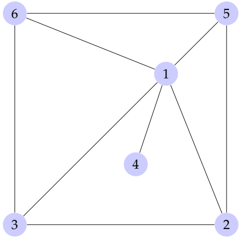


Similar construction with constraints $\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geqslant 0\}$

What is correlative sparsity?

$$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

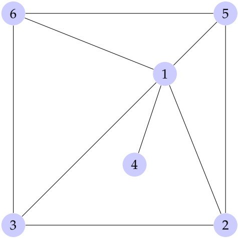
Chordal graph after adding edge $(3,5)$



What is correlative sparsity?

$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$

Chordal graph after adding edge $(3,5)$



maximal cliques $I_1 = \{1,4\}$ $I_2 = \{1,2,3,5\}$ $I_3 = \{1,3,5,6\}$

$f = f_1 + f_2 + f_3$ where f_k involves **only** variables in I_k

Let us index moment matrices and SOS with the cliques I_k

A sparse variant of Putinar's Positivstellensatz

Convergence of the Moment-SOS hierarchy is based on:

Theorem [Putinar '93] Positivstellensatz

If $\mathbf{X}$ contains a ball constraint $N - \sum_i x_i^2 \geqslant 0$ then

$$f > 0 \text{ on } \mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\} \implies f = \sigma_0 + \sum_j \sigma_j g_j \text{ with } \sigma_j \text{ SOS}$$

A sparse variant of Putinar's Positivstellensatz

Convergence of the Moment-SOS hierarchy is based on:

Theorem [Putinar '93] Positivstellensatz

If $\mathbf{X}$ contains a ball constraint $N - \sum_i x_i^2 \geqslant 0$ then

$$f > 0 \text{ on } \mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\} \implies f = \sigma_0 + \sum_j \sigma_j g_j \text{ with } \sigma_j \text{ SOS}$$

Theorem: Sparse Putinar's representation [Lasserre '06]

$f = \sum_k f_k$, f_k depends on $\mathbf{x}(I_k)$

$f > 0$ on $\mathbf{X}$

Each g_j depends on some I_k

RIP holds for (I_k)

ball constraints for each $\mathbf{x}(I_k)$

$\implies$

$$\boxed{f = \sum_k (\sigma_{0k} + \sum_{j \in J_k} \sigma_j g_j)}$$

SOS σ_{0k} "sees" vars in I_k

σ_j "sees" vars from g_j

A first key message

SUMS OF SQUARES PRESERVE SPARSITY

A proof of sparse Putinar's Positivstellensatz

Let $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$ be compact and $f = \sum_k f_k$, with f_k depends on $\mathbf{x}(I_k)$, and $f > 0$ on $\mathbf{X}$

A proof of sparse Putinar's Positivstellensatz

Let $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$ be compact and $f = \sum_k f_k$, with f_k depends on $\mathbf{x}(I_k)$, and $f > 0$ on $\mathbf{X}$

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geqslant 0 : j \in J_k\}$ = the subspace of $\mathbf{X}$ which only "sees" variables indexed by I_k

A proof of sparse Putinar's Positivstellensatz

Let $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$ be compact and $f = \sum_k f_k$, with f_k depends on $\mathbf{x}(I_k)$, and $f > 0$ on $\mathbf{X}$

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geqslant 0 : j \in J_k\}$ = the subspace of $\mathbf{X}$ which only "sees" variables indexed by I_k

Lemma [Grimm et al. '07]

If RIP holds for (I_k) then
$f = \sum_k h_k$, with h_k depends on $\mathbf{x}(I_k)$, and $h_k > 0$ on $\mathbf{X}_k$

A proof of sparse Putinar's Positivstellensatz

Let $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$ be compact and $f = \sum_k f_k$, with f_k depends on $\mathbf{x}(I_k)$, and $f > 0$ on $\mathbf{X}$

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geqslant 0 : j \in J_k\}$ = the subspace of $\mathbf{X}$ which only "sees" variables indexed by I_k

Lemma [Grimm et al. '07]

If RIP holds for (I_k) then
$f = \sum_k h_k$, with h_k depends on $\mathbf{x}(I_k)$, and $h_k > 0$ on $\mathbf{X}_k$

Prove this lemma by induction on the number of subsets I_k

A proof of sparse Putinar's Positivstellensatz

Let $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$ be compact and $f = \sum_k f_k$, with f_k depends on $\mathbf{x}(I_k)$, and $f > 0$ on $\mathbf{X}$

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geqslant 0 : j \in J_k\}$ = the subspace of $\mathbf{X}$ which only "sees" variables indexed by I_k

Lemma [Grimm et al. '07]

If RIP holds for (I_k) then
$f = \sum_k h_k$, with h_k depends on $\mathbf{x}(I_k)$, and $h_k > 0$ on $\mathbf{X}_k$

- Prove this lemma by induction on the number of subsets I_k

- Then apply Putinar to each h_k

Sparse moment matrices

For each subset I_k, submatrix of $\mathbf{M}_r(\mathbf{y})$ corresponding of rows & columns indexed by monomials in $\mathbf{x}(I_k)$

Sparse moment matrices

For each subset I_k, submatrix of $\mathbf{M}_r(\mathbf{y})$ corresponding of rows & columns indexed by monomials in $\mathbf{x}(I_k)$

$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$

Sparse moment matrices

For each subset I_k, submatrix of $\mathbf{M}_r(\mathbf{y})$ corresponding of rows & columns indexed by monomials in $\mathbf{x}(I_k)$

$$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

$I_1 = \{1, 4\} \implies$ monomials in x_1, x_4

$$\mathbf{M}_1(\mathbf{y}, I_1) = \begin{pmatrix} 1 & | & y_{1,0,0,0,0,0} & y_{0,0,0,1,0,0} \\ & - & - & - \\ y_{1,0,0,0,0,0} & | & y_{2,0,0,0,0,0} & y_{1,0,0,1,0,0} \\ y_{0,0,0,1,0,0} & | & y_{1,0,0,1,0,0} & y_{0,0,0,2,0,0} \end{pmatrix}$$

Sparse moment matrices

For each subset I_k, submatrix of $\mathbf{M}_r(\mathbf{y})$ corresponding of rows & columns indexed by monomials in $\mathbf{x}(I_k)$

$$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

$I_1 = \{1, 4\} \implies$ monomials in x_1, x_4

$$\mathbf{M}_1(\mathbf{y}, I_1) = \begin{pmatrix} 1 & | & y_{1,0,0,0,0,0} & y_{0,0,0,1,0,0} \\ & - & - & - \\ y_{1,0,0,0,0,0} & | & y_{2,0,0,0,0,0} & y_{1,0,0,1,0,0} \\ y_{0,0,0,1,0,0} & | & y_{1,0,0,1,0,0} & y_{0,0,0,2,0,0} \end{pmatrix}$$

same for each localizing matrix $\mathbf{M}_r(g_j\mathbf{y})$

Sparse primal-dual Moment-SOS hierarchy

$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$ with $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$

Dense Moment-SOS hierarchy

(Moment)		(SOS)
$\inf \sum_{\alpha} f_\alpha \, y_\alpha$	$=$	$\sup \; \lambda$
s.t. $\mathbf{M}_r(\mathbf{y}) \succcurlyeq 0$		s.t. $\lambda \in \mathbb{R}$
$\mathbf{M}_{r-r_j}(g_j \, \mathbf{y}) \succcurlyeq 0$		$f - \lambda = \sigma_0 + \sum_j \sigma_j g_j$
$y_0 = 1$		

Sparse primal-dual Moment-SOS hierarchy

$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$ with $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$
$f = \sum_k f_k$, with f_k depends on $\mathbf{x}(I_k)$
Each g_j depends on some I_k

Sparse Moment-SOS hierarchy

(Moment) = (SOS)

$$\begin{aligned} \inf \quad & \sum_\alpha f_\alpha y_\alpha \\ \text{s.t.} \quad & \mathbf{M}_r(\mathbf{y}, I_k) \succcurlyeq 0 \\ & \mathbf{M}_{r-r_j}(g_j\, \mathbf{y}, I_k) \succcurlyeq 0\,, j \in J_k, \forall k \\ & y_0 = 1 \end{aligned}
\quad = \quad
\begin{aligned} \sup \quad & \lambda \\ \text{s.t.} \quad & \lambda \in \mathbb{R} \\ & f - \lambda = \sum_k \Big(\sigma_{k0} + \sum_{j \in J_k} \sigma_j g_j\Big) \end{aligned}$$

Sparse primal-dual Moment-SOS hierarchy

$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$ with $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$
$f = \sum_k f_k$, with f_k depends on $\mathbf{x}(I_k)$
Each g_j depends on some I_k

Sparse Moment-SOS hierarchy

(Moment) (SOS)

$$
\begin{array}{llcll}
\inf & \sum_{\alpha} f_\alpha y_\alpha & = & \sup & \lambda \\
\text{s.t.} & \mathbf{M}_r(\mathbf{y}, I_k) \succcurlyeq 0 & & \text{s.t.} & \lambda \in \mathbb{R} \\
 & \mathbf{M}_{r-r_j}(g_j\,\mathbf{y}, I_k) \succcurlyeq 0\,, j \in J_k, \forall k & & & f - \lambda = \sum_k (\sigma_{k0} + \sum_{j \in J_k} \sigma_j g_j) \\
 & y_0 = 1 & & &
\end{array}
$$

RIP holds for (I_k) + ball constraints for each $\mathbf{x}(I_k) \implies$ Primal and dual optimal value converge to $f_{\min}$ by sparse Putinar

Computational cost

$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$ with $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0, j \leqslant m\}$
$\tau = \max\{|I_1|, \ldots, |I_p|\}$

Sparse Moment-SOS hierarchy

(Moment)		=	(SOS)	
inf	$\sum_\alpha f_\alpha y_\alpha$	=	sup	λ
s.t.	$\mathbf{M}_r(\mathbf{y}, I_k) \succcurlyeq 0$		s.t.	$\lambda \in \mathbb{R}$
	$\mathbf{M}_{r-r_j}(g_j\, \mathbf{y}, I_k) \succcurlyeq 0\,, j \in J_k, \forall k$			$f - \lambda = \sum_k (\sigma_{k0} + \sum_{j \in J_k} \sigma_j g_j)$
	$y_0 = 1$			

Computational cost

$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$ with $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0, j \leqslant m\}$
$\tau = \max\{|I_1|, \ldots, |I_p|\}$

Sparse Moment-SOS hierarchy

(Moment) (SOS)

$$\inf \sum_\alpha f_\alpha y_\alpha \quad = \quad \sup \lambda$$

s.t. $\mathbf{M}_r(\mathbf{y}, I_k) \succcurlyeq 0$ s.t. $\lambda \in \mathbb{R}$

$\mathbf{M}_{r-r_j}(g_j \, \mathbf{y}, I_k) \succcurlyeq 0 \, , j \in J_k, \forall k$ $\quad f - \lambda = \sum_k (\sigma_{k0} + \sum_{j \in J_k} \sigma_j g_j)$

$y_0 = 1$

$(m + p)$ SOS in at most τ vars of degree $\leqslant 2r$

Computational cost

$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$ with $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0, j \leqslant m\}$
$\tau = \max\{|I_1|, \ldots, |I_p|\}$

Sparse Moment-SOS hierarchy

(Moment) | (SOS)

$$\inf \sum_\alpha f_\alpha y_\alpha = \sup \lambda$$

(Moment) s.t. $\mathbf{M}_r(\mathbf{y}, I_k) \succcurlyeq 0$

$\mathbf{M}_{r-r_j}(g_j\, \mathbf{y}, I_k) \succcurlyeq 0\,, j \in J_k, \forall k$

$y_0 = 1$

(SOS) s.t. $\lambda \in \mathbb{R}$

$f - \lambda = \sum_k (\sigma_{k0} + \sum_{j \in J_k} \sigma_j g_j)$

$(m+p)$ SOS in at most τ vars of degree $\leqslant 2r$

$(m+p)\, \mathcal{O}\left(r^{2\tau}\right)$ SDP vars

Computational cost

$f_{\min} = \inf_{\mathbf{x}\in\mathbf{X}} f(\mathbf{x})$ with $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0, j \leqslant m\}$
$\tau = \max\{|I_1|, \ldots, |I_p|\}$

Sparse Moment-SOS hierarchy

(Moment) (SOS)

$$\begin{aligned}
\inf \quad & \sum_\alpha f_\alpha \, y_\alpha & = \quad & \sup \quad \lambda \\
\text{s.t.} \quad & \mathbf{M}_r(\mathbf{y}, I_k) \succcurlyeq 0 & & \text{s.t.} \quad \lambda \in \mathbb{R} \\
& \mathbf{M}_{r-r_j}(g_j \, \mathbf{y}, I_k) \succcurlyeq 0 \,, j \in J_k, \forall k & & f - \lambda = \sum_k (\sigma_{k0} + \sum_{j \in J_k} \sigma_j g_j) \\
& y_0 = 1 & &
\end{aligned}$$

$(m + p)$ SOS in at most τ vars of degree $\leqslant 2r$
$(m+p)\,\mathcal{O}\left(r^{2\tau}\right)$ SDP vars vs $(m+1)\,\mathcal{O}\left(r^{2n}\right)$ in the dense SDP

Sparse linear program over measures

In the dense setting:

$$\begin{aligned} f_{\min} = \inf_{\mu} \quad & \int_{\mathbf{X}} f \, d\mu \\ \text{s.t.} \quad & \mu \in \mathcal{M}_+(\mathbf{X}) \end{aligned}$$

Sparse linear program over measures

In the dense setting:

$$\begin{aligned} f_{\min} = \inf_{\mu} \quad & \int_{\mathbf{X}} f \, d\mu \\ \text{s.t.} \quad & \mu \in \mathcal{M}_+(\mathbf{X}) \end{aligned}$$

In the sparse setting:

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geqslant 0 : j \in J_k\}$ = the subspace of $\mathbf{X}$ which only "sees" variables indexed by I_k

Sparse linear program over measures

In the dense setting:

$$\begin{aligned} f_{\min} = \inf_{\mu} \quad & \int_{\mathbf{X}} f \, d\mu \\ \text{s.t.} \quad & \mu \in \mathcal{M}_+(\mathbf{X}) \end{aligned}$$

In the sparse setting:

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geqslant 0 : j \in J_k\}$ = the subspace of $\mathbf{X}$ which only "sees" variables indexed by I_k

$\mathbf{X}_{kj}$ = only "sees" variables indexed by $I_k \cap I_j$

Sparse linear program over measures

In the dense setting:

$$f_{\min} = \inf_{\mu} \int_{\mathbf{X}} f \, d\mu \quad \text{s.t.} \quad \mu \in \mathcal{M}_+(\mathbf{X})$$

In the sparse setting:

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geqslant 0 : j \in J_k\}$ = the subspace of $\mathbf{X}$ which only "sees" variables indexed by I_k

$\mathbf{X}_{kj}$ = only "sees" variables indexed by $I_k \cap I_j$

💡 one measure μ_k for each $I_k \to$ marginals $\pi_{kj}\mu_k$ on $\mathcal{M}_+(\mathbf{X}_{kj})$

Sparse linear program over measures

In the dense setting:

$$
\begin{aligned}
f_{\min} = \inf_{\mu} \quad & \int_{\mathbf{X}} f \, d\mu \\
\text{s.t.} \quad & \mu \in \mathcal{M}_+(\mathbf{X})
\end{aligned}
$$

In the sparse setting:

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geqslant 0 : j \in J_k\}$ = the subspace of $\mathbf{X}$ which only "sees" variables indexed by I_k

$\mathbf{X}_{kj}$ = only "sees" variables indexed by $I_k \cap I_j$

one measure μ_k for each $I_k \to$ marginals $\pi_{kj}\mu_k$ on $\mathcal{M}_+(\mathbf{X}_{kj})$

Sparse moment SDPs relax the sparse LP over measures:

$$
\begin{aligned}
f_{\text{cs}} = \inf_{\mu_k} \quad & \sum_k \int_{\mathbf{X}_k} f_k \, d\mu_k \\
\text{s.t.} \quad & \pi_{jk}\mu_j = \pi_{kj}\mu_k \,, \quad \mu_k \in \mathcal{M}_+(\mathbf{X}_k)
\end{aligned}
$$

The dual of sparse Putinar's Positivstellensatz

Theorem [Lasserre '06]

$$\text{RIP holds for } (I_k) \implies f_{\min} = f_{\text{cs}} = \inf_{\mu_k} \sum_k \int_{\mathbf{X}_k} f_k \, d\mu_k$$
$$\text{s.t. } \pi_{jk}\mu_j = \pi_{kj}\mu_k$$
$$\mu_k \in \mathcal{M}_+(\mathbf{X}_k)$$

The dual of sparse Putinar's Positivstellensatz

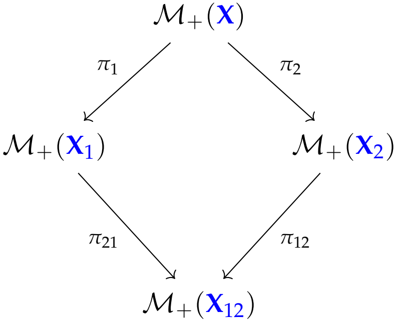
Theorem [Lasserre '06]

RIP holds for $(I_k) \implies f_{\min} = f_{\text{cs}} = \inf_{\mu_k} \sum_k \int_{\mathbf{X}_k} f_k \, d\mu_k$

s.t. $\pi_{jk}\mu_j = \pi_{kj}\mu_k$

$\mu_k \in \mathcal{M}_+(\mathbf{X}_k)$

Proof: there exists $\mu \in \mathcal{M}_+(\mathbf{X})$ with marginal μ_k on $\mathbf{X}_k$



A first (dual) key message

THE MEASURE LP PRESERVES SPARSITY

Extracting minimizers: the dense case

Let $r_{\min}$ be the minimal relaxation order.

Extracting minimizers: the dense case

Let $r_{\min}$ be the minimal relaxation order.

Theorem: dense extraction [Lasserre & Henrion '05]

Assume that the moment SDP has an optimal solution $\mathbf{y}$ with cost f^r and

$$\operatorname{rank} \mathbf{M}_{r'}(\mathbf{y}) = \operatorname{rank} \mathbf{M}_{r'-r_{\min}}(\mathbf{y}) \text{ for some } r' \leqslant r\,.$$

Extracting minimizers: the dense case

Let $r_{\min}$ be the minimal relaxation order.

Theorem: dense extraction [Lasserre & Henrion '05]

Assume that the moment SDP has an optimal solution $\mathbf{y}$ with cost f^r and

$$\operatorname{rank} \mathbf{M}_{r'}(\mathbf{y}) = \operatorname{rank} \mathbf{M}_{r'-r_{\min}}(\mathbf{y}) \text{ for some } r' \leqslant r\,.$$

Then $f^r = f_{\min}$ and the LP over measures has an optimal solution $\mu \in \mathcal{M}_+(\mathbf{X})$ supported on $t = \operatorname{rank} \mathbf{M}_{r'}(\mathbf{y})$ points of $\mathbf{X}$.

Extracting minimizers: the dense case

Let $r_{\min}$ be the minimal relaxation order.

Theorem: dense extraction [Lasserre & Henrion '05]

Assume that the moment SDP has an optimal solution $\mathbf{y}$ with cost f^r and

$$\operatorname{rank} \mathbf{M}_{r'}(\mathbf{y}) = \operatorname{rank} \mathbf{M}_{r'-r_{\min}}(\mathbf{y}) \text{ for some } r' \leqslant r\,.$$

Then $f^r = f_{\min}$ and the LP over measures has an optimal solution $\mu \in \mathcal{M}_+(\mathbf{X})$ supported on $t = \operatorname{rank} \mathbf{M}_{r'}(\mathbf{y})$ points of $\mathbf{X}$.

Extraction possible with the `Gloptipoly` software

Extracting minimizers: the sparse case

$a_k = \max_{j \in J_k} r_j$ = max half degree of g_j depending on I_k.

Extracting minimizers: the sparse case

$a_k = \max_{j \in J_k} r_j$ = max half degree of g_j depending on I_k.

Theorem: sparse extraction [Lasserre '06]

Assume that the sparse moment SDP has an optimal solution $\mathbf{y}$ with cost f^r_{cs} and

$$\operatorname{rank} \mathbf{M}_r(\mathbf{y}, I_k) = \operatorname{rank} \mathbf{M}_{r-a_k}(\mathbf{y}, I_k)$$
$$\operatorname{rank} \mathbf{M}_r(\mathbf{y}, I_k \cap I_j) = 1$$

Extracting minimizers: the sparse case

$a_k = \max_{j \in J_k} r_j$ = max half degree of g_j depending on I_k.

Theorem: sparse extraction [Lasserre '06]

Assume that the sparse moment SDP has an optimal solution $\mathbf{y}$ with cost f^r_{cs} and

$$\operatorname{rank} \mathbf{M}_r(\mathbf{y}, I_k) = \operatorname{rank} \mathbf{M}_{r-a_k}(\mathbf{y}, I_k)$$
$$\operatorname{rank} \mathbf{M}_r(\mathbf{y}, I_k \cap I_j) = 1$$

Then $f^r_{\text{cs}} = f_{\min} = f_{\text{cs}}$ & sparse measure LP has optimal solution $\mu_k \in \mathcal{M}_+(\mathbf{X}_k)$ supported on $t_k = \operatorname{rank} \mathbf{M}_r(\mathbf{y}, I_k)$ points of $\mathbf{X}_k$.

Extracting minimizers: the sparse case

$a_k = \max_{j \in J_k} r_j$ = max half degree of g_j depending on I_k.

Theorem: sparse extraction [Lasserre '06]

Assume that the sparse moment SDP has an optimal solution $\mathbf{y}$ with cost f^r_{cs} and

$$\operatorname{rank} \mathbf{M}_r(\mathbf{y}, I_k) = \operatorname{rank} \mathbf{M}_{r-a_k}(\mathbf{y}, I_k)$$
$$\operatorname{rank} \mathbf{M}_r(\mathbf{y}, I_k \cap I_j) = 1$$

Then $f^r_{\text{cs}} = f_{\min} = f_{\text{cs}}$ & sparse measure LP has optimal solution $\mu_k \in \mathcal{M}_+(\mathbf{X}_k)$ supported on $t_k = \operatorname{rank} \mathbf{M}_r(\mathbf{y}, I_k)$ points of $\mathbf{X}_k$.

- RIP is not required!
- Extract $\mathbf{x}(k)$ from $\mathbf{M}_r(\mathbf{y}, I_k) \implies$ minimizer $\mathbf{x}$ with $(x_i)_{i \in I_k} = \mathbf{x}(k)$

Application to rational functions

$$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} \sum_i \frac{p_i(\mathbf{x})}{q_i(\mathbf{x})}, \quad q_i > 0 \text{ on } \mathbf{X}, \quad p_i, q_i \text{ depends only on } I_i$$

Application to rational functions

$$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} \sum_i \frac{p_i(\mathbf{x})}{q_i(\mathbf{x})}, \quad q_i > 0 \text{ on } \mathbf{X}, \quad p_i, q_i \text{ depends only on } I_i$$

Theorem: dense measure LP [Bugarin et al. '16]

$$\begin{aligned} f_{\min} = \inf_{\mu_i \in \mathcal{M}_+(\mathbf{X})} \quad & \sum_i \int_{\mathbf{X}} p_i \, d\mu_i \\ \text{s.t.} \quad & \int_{\mathbf{X}} \mathbf{x}^\alpha q_i \, d\mu_i = \int_{\mathbf{X}} \mathbf{x}^\alpha q_1 \, d\mu_1, \ \alpha \in \mathbb{N}^n \\ & \int_{\mathbf{X}} q_1 \, d\mu_1 = 1 \end{aligned}$$

Application to rational functions

$$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} \sum_i \frac{p_i(\mathbf{x})}{q_i(\mathbf{x})}, \quad q_i > 0 \text{ on } \mathbf{X}, \quad p_i, q_i \text{ depends only on } I_i$$

Theorem: dense measure LP [Bugarin et al. '16]

$$\begin{aligned} f_{\min} = \inf_{\mu_i \in \mathcal{M}_+(\mathbf{X})} \quad & \sum_i \int_{\mathbf{X}} p_i \, d\mu_i \\ \text{s.t.} \quad & \int_{\mathbf{X}} \mathbf{x}^\alpha \, q_i \, d\mu_i = \int_{\mathbf{X}} \mathbf{x}^\alpha \, q_1 \, d\mu_1 \,, \; \alpha \in \mathbb{N}^n \\ & \int_{\mathbf{X}} q_1 \, d\mu_1 = 1 \end{aligned}$$

Theorem: sparse measure LP [Bugarin et al. '16]

$$\begin{aligned} f_{\min} = f_{\text{cs}} = \inf_{\mu_i \in \mathcal{M}_+(\mathbf{X}_i)} \quad & \sum_i \int_{\mathbf{X}_i} p_i \, d\mu_i \\ \text{s.t.} \quad & \pi_{ij}(q_i d\mu_i) = \pi_{ji}(q_j d\mu_j) \\ & \int_{\mathbf{X}_i} q_i \, d\mu_i = 1 \end{aligned}$$

Application to roundoff errors

[Magron Constantinides Donaldson '17]

Exact $f(\mathbf{x}) = x_1 x_2 + x_3 x_4$

Application to roundoff errors

[Magron Constantinides Donaldson '17]

Exact $f(\mathbf{x}) = x_1 x_2 + x_3 x_4$
Floating-point $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1 x_2 (1 + e_1) + x_3 x_4 (1 + e_2)](1 + e_3)$

Application to roundoff errors

[Magron Constantinides Donaldson '17]

Exact $f(\mathbf{x}) = x_1x_2 + x_3x_4$
Floating-point $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1x_2(1 + e_1) + x_3x_4(1 + e_2)](1 + e_3)$
$\mathbf{x} \in \mathbf{X}$, $\quad \mid e_i \mid \leqslant 2^{-\delta} \quad \delta = 24$ (single) or 53 (double)

Application to roundoff errors

[Magron Constantinides Donaldson '17]

Exact $f(\mathbf{x}) = x_1 x_2 + x_3 x_4$
Floating-point $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1 x_2 (1 + e_1) + x_3 x_4 (1 + e_2)](1 + e_3)$
$\mathbf{x} \in \mathbf{X}, \quad | e_i | \leqslant 2^{-\delta} \quad \delta = 24$ (single) or 53 (double)

1: Error $f(\mathbf{x}) - \hat{f}(\mathbf{x}, \mathbf{e}) = \ell(\mathbf{x}, \mathbf{e}) + h(\mathbf{x}, \mathbf{e})$, ℓ **linear** in $\mathbf{e}$

Application to roundoff errors

[Magron Constantinides Donaldson '17]

Exact $f(\mathbf{x}) = x_1 x_2 + x_3 x_4$
Floating-point $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1 x_2 (1 + e_1) + x_3 x_4 (1 + e_2)](1 + e_3)$
$\mathbf{x} \in \mathbf{X}$, $\mid e_i \mid \leqslant 2^{-\delta}$ $\delta = 24$ (single) or 53 (double)

1: Error $f(\mathbf{x}) - \hat{f}(\mathbf{x}, \mathbf{e}) = \ell(\mathbf{x}, \mathbf{e}) + h(\mathbf{x}, \mathbf{e})$, ℓ **linear** in $\mathbf{e}$

2: Bound $h(\mathbf{x}, \mathbf{e})$ with interval arithmetic

Application to roundoff errors

[Magron Constantinides Donaldson '17]

Exact $f(\mathbf{x}) = x_1 x_2 + x_3 x_4$
Floating-point $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1 x_2 (1 + e_1) + x_3 x_4 (1 + e_2)](1 + e_3)$
$\mathbf{x} \in \mathbf{X}, \quad | e_i | \leqslant 2^{-\delta} \quad \delta = 24$ (single) or 53 (double)

1: Error $f(\mathbf{x}) - \hat{f}(\mathbf{x}, \mathbf{e}) = \ell(\mathbf{x}, \mathbf{e}) + h(\mathbf{x}, \mathbf{e})$, ℓ **linear** in $\mathbf{e}$

2: Bound $h(\mathbf{x}, \mathbf{e})$ with interval arithmetic

3: Bound $\ell(\mathbf{x}, \mathbf{e})$ with **SPARSE SUMS OF SQUARES**

💡 $I_k \to \{\mathbf{x}, e_k\} \implies \boxed{m(n+1)^{2r} \text{ instead of } (n+m)^{2r}}$ SDP vars

Application to roundoff errors

$$f = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

$$\mathbf{x} \in [4.00, 6.36]^6\,, \quad \mathbf{e} \in [-\epsilon, \epsilon]^{15}\,, \quad \epsilon = 2^{-53}$$

Dense SDP: $\binom{6+15+4}{6+15} = 12650$ variables $\leadsto$ **Out of memory**

Application to roundoff errors

$$f = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

$$\mathbf{x} \in [4.00, 6.36]^6 , \quad \mathbf{e} \in [-\epsilon, \epsilon]^{15} , \quad \epsilon = 2^{-53}$$

Dense SDP: $\binom{6+15+4}{6+15} = 12650$ variables $\leadsto$ **Out of memory**

Sparse SDP `Real2Float` tool: $15\binom{6+1+4}{6+1} = 4950 \leadsto 759\epsilon$

Application to roundoff errors

$$f = x_2 x_5 + x_3 x_6 - x_2 x_3 - x_5 x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

$$\mathbf{x} \in [4.00, 6.36]^6 , \quad \mathbf{e} \in [-\epsilon, \epsilon]^{15} , \quad \epsilon = 2^{-53}$$

Dense SDP: $\binom{6+15+4}{6+15} = 12650$ variables $\rightsquigarrow$ **Out of memory**

Sparse SDP `Real2Float` tool: $15\binom{6+1+4}{6+1} = 4950 \rightsquigarrow 759\epsilon$

Interval arithmetic: 922ϵ (10 $\times$ less CPU)

Application to roundoff errors

$$f = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

$$\mathbf{x} \in [4.00, 6.36]^6\,, \quad \mathbf{e} \in [-\epsilon, \epsilon]^{15}\,, \quad \epsilon = 2^{-53}$$

Dense SDP: $\binom{6+15+4}{6+15} = 12650$ variables $\rightsquigarrow$ **Out of memory**

Sparse SDP `Real2Float` tool: $15\binom{6+1+4}{6+1} = 4950 \rightsquigarrow 759\epsilon$

Interval arithmetic: 922ϵ (10 $\times$ less CPU)

Symbolic Taylor `FPTaylor` tool: 721ϵ (21 $\times$ more CPU)

Application to roundoff errors

$$f = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

$$\mathbf{x} \in [4.00, 6.36]^6, \quad \mathbf{e} \in [-\epsilon, \epsilon]^{15}, \quad \epsilon = 2^{-53}$$

Dense SDP: $\binom{6+15+4}{6+15} = 12650$ variables $\leadsto$ **Out of memory**

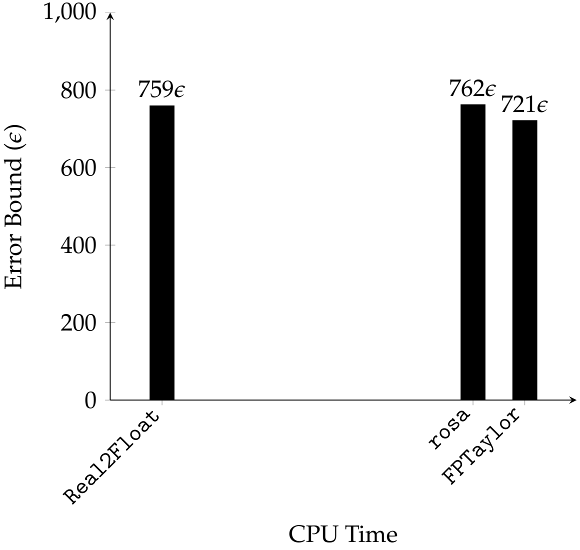
Sparse SDP `Real2Float` tool: $15\binom{6+1+4}{6+1} = 4950 \leadsto 759\epsilon$

Interval arithmetic: 922ϵ (10 $\times$ less CPU)

Symbolic Taylor `FPTaylor` tool: 721ϵ (21 $\times$ more CPU)

SMT-based `rosa` tool: 762ϵ (19 $\times$ more CPU)

Application to roundoff errors

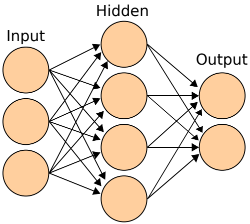


Application to deep learning

[SIAM News March ’21]

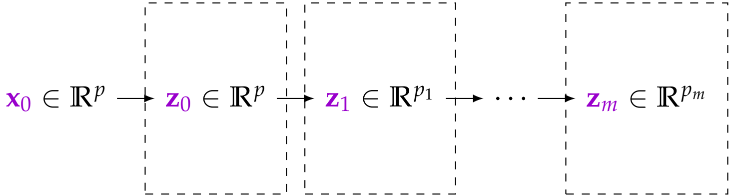
“Yet DL has an Achilles’ heel. Current implementations can be highly unstable, meaning that a certain small perturbation to the input of a trained neural network can cause substantial change in its output. This phenomenon is both a nuisance and a major concern for the safety and robustness of DL-based systems in critical applications—like healthcare—where reliable computations are essential”

Application to deep learning



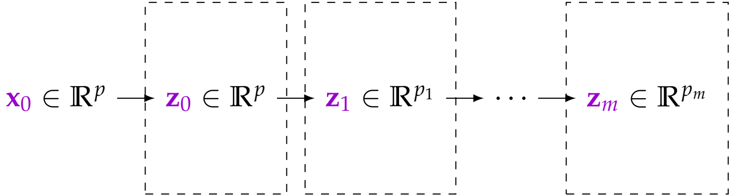
- Applications: WGAN, certification
- Existing works: [Lattore et al.'18] based on linear programming (LP)
- Network setting: K-classifier, **ReLU network**, $1 + m$ layers (1 input layer + m hidden layer), $\mathbf{A}_i$ weights, $\mathbf{b}_i$ biases
- Score of label $k \leqslant K$ = $\mathbf{c}_k{}^T \mathbf{x}_m$ with last activation vector $\mathbf{c}_k$

Application to deep learning



$\mathbf{z}_i = \mathbf{A}_i \mathbf{x}_{i-1} + \mathbf{b}_i \qquad \mathbf{x}_{i-1} = \mathrm{ReLU}(\mathbf{z}_{i-1})$

Application to deep learning



$\mathbf{z}_i = \mathbf{A}_i \mathbf{x}_{i-1} + \mathbf{b}_i \qquad \mathbf{x}_{i-1} = \mathrm{ReLU}(\mathbf{z}_{i-1})$

LIPSCHITZ CONSTANT:

$$L_f^{||\cdot||} = \inf\{L : \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}, |f(\mathbf{x}) - f(\mathbf{y})| \leq L||\mathbf{x} - \mathbf{y}||\}$$

Application to deep learning

$$\mathbf{x}_0 \in \mathbb{R}^p \longrightarrow \mathbf{z}_0 \in \mathbb{R}^p \longrightarrow \mathbf{z}_1 \in \mathbb{R}^{p_1} \longrightarrow \cdots \longrightarrow \mathbf{z}_m \in \mathbb{R}^{p_m}$$

$$\mathbf{z}_i = \mathbf{A}_i \mathbf{x}_{i-1} + \mathbf{b}_i \qquad \mathbf{x}_{i-1} = \mathrm{ReLU}(\mathbf{z}_{i-1})$$

LIPSCHITZ CONSTANT:

$$\begin{aligned} L_f^{||\cdot||} &= \inf\{L : \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}, |f(\mathbf{x}) - f(\mathbf{y})| \leq L||\mathbf{x} - \mathbf{y}||\} \\ &= \sup\{||\nabla f(\mathbf{x})||_* : \mathbf{x} \in \mathcal{X}\} \\ &= \sup\{\mathbf{t}^T \nabla f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}, ||\mathbf{t}|| \leq 1\} \end{aligned}$$

Application to deep learning

$$\mathbf{x}_0 \in \mathbb{R}^p \longrightarrow \mathbf{z}_0 \in \mathbb{R}^p \longrightarrow \mathbf{z}_1 \in \mathbb{R}^{p_1} \longrightarrow \cdots \longrightarrow \mathbf{z}_m \in \mathbb{R}^{p_m}$$

$$\mathbf{z}_i = \mathbf{A}_i \mathbf{x}_{i-1} + \mathbf{b}_i \qquad \mathbf{x}_{i-1} = \mathrm{ReLU}(\mathbf{z}_{i-1})$$

LIPSCHITZ CONSTANT:

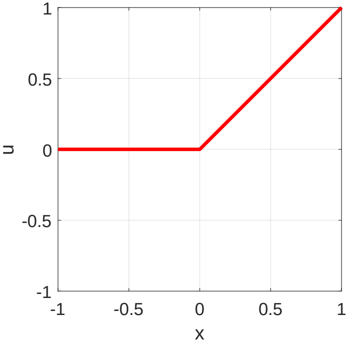
$$\begin{aligned} L_f^{||\cdot||} &= \inf\{L : \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}, |f(\mathbf{x}) - f(\mathbf{y})| \leq L||\mathbf{x} - \mathbf{y}||\} \\ &= \sup\{||\nabla f(\mathbf{x})||_* : \mathbf{x} \in \mathcal{X}\} \\ &= \sup\{\mathbf{t}^T \nabla f(\mathbf{x}) : \mathbf{x} \in \mathcal{X}, ||\mathbf{t}|| \leq 1\} \end{aligned}$$

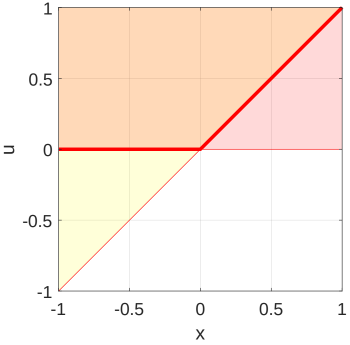
GRADIENT for a fixed label k:

$$\nabla f(\mathbf{x}_0) = \left(\prod_{i=1}^{m} \mathbf{A}_i{}^T \mathrm{diag}\,(\mathrm{ReLU}'(\mathbf{z}_i)) \right) \mathbf{c}_k$$

Correlative sparsity in POP

ReLU (left) & its semialgebraicity (right)

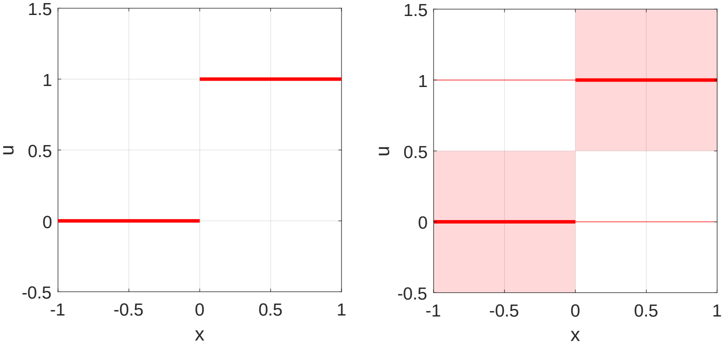




$$u = \max\{x, 0\} \qquad\qquad u(u - x) = 0, u \geq x, u \geq 0$$

Correlative sparsity in POP

ReLU' (left) & its semialgebraicity (right)



$u = \mathbf{1}_{\{x \geq 0\}}$ $\qquad$ $u(u-1) = 0, (u - \frac{1}{2})x \geq 0$

Application to deep learning

Local Lipschitz constant: $\mathbf{x}_0 \in$ ball of center $\bar{\mathbf{x}}_0$ and radius ε

Application to deep learning

Local Lipschitz constant: $\mathbf{x}_0 \in$ ball of center $\bar{\mathbf{x}}_0$ and radius ε

One single hidden layer ($m = 1$):

$$\sup_{\mathbf{x},\mathbf{u},\mathbf{z},\mathbf{t}} \mathbf{t}^T \mathbf{A}^T \text{diag}(\mathbf{u})\mathbf{c}$$

$$\text{s.t.} \begin{cases} (\mathbf{z} - \mathbf{A}\mathbf{x} - \mathbf{b})^2 = 0 \\ \mathbf{t}^2 \leq 1, (\mathbf{x} - \bar{\mathbf{x}}_0 + \varepsilon)(\mathbf{x} - \bar{\mathbf{x}}_0 - \varepsilon) \leq 0 \\ \mathbf{u}(\mathbf{u} - 1) = 0, (\mathbf{u} - 1/2)\mathbf{z} \geq 0 \end{cases}$$

Application to deep learning

Local Lipschitz constant: $\mathbf{x}_0 \in$ ball of center $\bar{\mathbf{x}}_0$ and radius ε

One single hidden layer ($m = 1$):

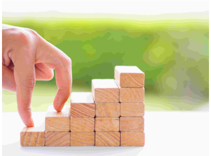
$$\sup_{\mathbf{x},\mathbf{u},\mathbf{z},\mathbf{t}} \mathbf{t}^T \mathbf{A}^T \text{diag}\,(\mathbf{u})\mathbf{c}$$

$$\text{s.t.} \begin{cases} (\mathbf{z} - \mathbf{A}\mathbf{x} - \mathbf{b})^2 = 0 \\ \mathbf{t}^2 \leq 1, (\mathbf{x} - \bar{\mathbf{x}}_0 + \varepsilon)(\mathbf{x} - \bar{\mathbf{x}}_0 - \varepsilon) \leq 0 \\ \mathbf{u}(\mathbf{u} - 1) = 0, (\mathbf{u} - 1/2)\mathbf{z} \geq 0 \end{cases}$$

"CHEAP" and "TIGHT" upper bound?

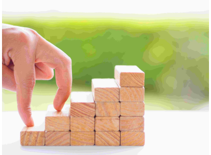
Our “heuristic relaxation” method: HR-2

💡 Go between 1ST & 2ND stair in SPARSE hierarchy



Our "heuristic relaxation" method: HR-2

Go between 1ST & 2ND stair in SPARSE hierarchy

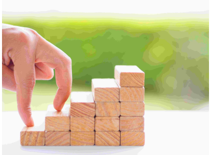


$$\sup_{\mathbf{x},\mathbf{u},\mathbf{z},\mathbf{t}} \ \mathbf{t}^T \mathbf{A}^T \mathrm{diag}\,(\mathbf{u})\mathbf{c}$$

$$\text{s.t.} \begin{cases} (\mathbf{z} - \mathbf{A}\mathbf{x} - \mathbf{b})^2 = 0 \\ \mathbf{t}^2 \leq 1, (\mathbf{x} - \bar{\mathbf{x}}_0 + \varepsilon)(\mathbf{x} - \bar{\mathbf{x}}_0 - \varepsilon) \leq 0 \\ \mathbf{u}(\mathbf{u} - 1) = 0, (\mathbf{u} - 1/2)\mathbf{z} \geq 0 \end{cases}$$

Our “heuristic relaxation” method: HR-2

Go between 1ST & 2ND stair in SPARSE hierarchy



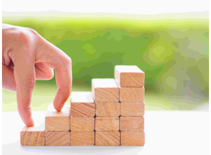
$$\sup_{\mathbf{x},\mathbf{u},\mathbf{z},\mathbf{t}} \mathbf{t}^T \mathbf{A}^T \text{diag}\,(\mathbf{u})\mathbf{c}$$

$$\text{s.t.} \begin{cases} (\mathbf{z} - \mathbf{A}\mathbf{x} - \mathbf{b})^2 = 0 \\ \mathbf{t}^2 \leq 1, (\mathbf{x} - \bar{\mathbf{x}}_0 + \varepsilon)(\mathbf{x} - \bar{\mathbf{x}}_0 - \varepsilon) \leq 0 \\ \mathbf{u}(\mathbf{u} - 1) = 0, (\mathbf{u} - 1/2)\mathbf{z} \geq 0 \end{cases}$$

Pick SDP variables for products in $\{x, t\}$, $\{u, z\}$ up to deg 4

Our "heuristic relaxation" method: HR-2

Go between 1ST & 2ND stair in SPARSE hierarchy



$$\sup_{\mathbf{x},\mathbf{u},\mathbf{z},\mathbf{t}} \mathbf{t}^T \mathbf{A}^T \text{diag}(\mathbf{u})\mathbf{c}$$

$$\text{s.t.} \begin{cases} (\mathbf{z} - \mathbf{A}\mathbf{x} - \mathbf{b})^2 = 0 \\ \mathbf{t}^2 \leq 1, (\mathbf{x} - \bar{\mathbf{x}}_0 + \varepsilon)(\mathbf{x} - \bar{\mathbf{x}}_0 - \varepsilon) \leq 0 \\ \mathbf{u}(\mathbf{u} - 1) = 0, (\mathbf{u} - 1/2)\mathbf{z} \geq 0 \end{cases}$$

Pick SDP variables for products in $\{x, t\}$, $\{u, z\}$ up to deg 4

Pick SDP variables for products in $\{x, z\}$, $\{t, u\}$ up to deg 2

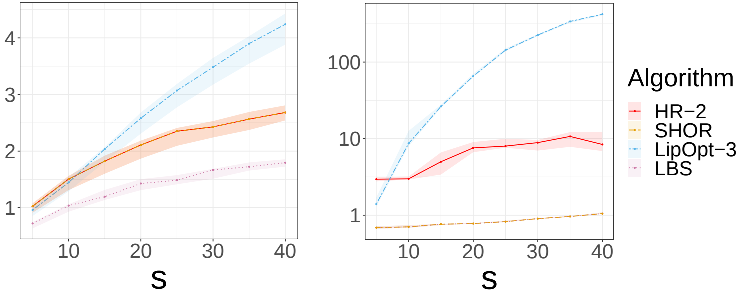
HR-2 on random $(80, 80)$ networks

Weight matrix $\mathbf{A}$ with band structure of width **s**

SHOR: Shor's relaxation given by 1ST stair in the hierarchy

LipOpt-3: LP based method

LBS: lower bound given by 10^4 random samples



Upper bound

Time

Application to SOS of bounded degrees

Theorem: sparse BSOS representation [Weisser et al. '18]

If $0 \leqslant g_j \leqslant 1$ on $\mathbf{X}$, $f > 0$ on $\mathbf{X}$ & RIP holds for (I_k) then

$$f = \sum_k \left(\sigma_k + \sum_{\alpha,\beta} c_{k,\alpha\beta} \prod_{j \in J_k} {g_j}^{\alpha_j} (1 - g_j)^{\beta_j} \right) ,$$

with σ_k SOS of degree $\leqslant 2r$

Application to sparse positive definite forms

Theorem: [Reznick '95] Positivstellensatz

pd form $f \implies \boxed{f = \frac{\sigma}{\|\mathbf{x}\|_2^{2r}}}$ with σ SOS, $r \in \mathbb{N}$

Application to sparse positive definite forms

Theorem: [Reznick '95] Positivstellensatz

pd form $f \implies \boxed{f = \frac{\sigma}{\|\mathbf{x}\|_2^{2r}}}$ with σ SOS, $r \in \mathbb{N}$

Sparse $f \sum_k f_k$, with f_k only depends on I_k

RUNNING INTERSECTION PROPERTY (RIP)

$$\boxed{\forall k \quad \underbrace{I_k \cap \bigcup_{j<k} I_j}_{\hat{I}_k} \subseteq I_{s_k} \quad \text{for some } s_k < k}$$

Application to sparse positive definite forms

Theorem: [Reznick '95] Positivstellensatz

pd form $f \implies \boxed{f = \frac{\sigma}{\|\mathbf{x}\|_2^{2r}}}$ with σ SOS, $r \in \mathbb{N}$

Sparse $f \sum_k f_k$, with f_k only depends on I_k

RUNNING INTERSECTION PROPERTY (RIP)

$$\boxed{\forall k \quad \underbrace{I_k \cap \bigcup_{j<k} I_j}_{\hat{I}_k} \subseteq I_{s_k} \quad \text{for some } s_k < k}$$

Theorem: sparse Reznick [Mai Lasserre Magron '20]

RIP $\implies \boxed{f = \sum_k \frac{\sigma_k}{{H_k}^r}}$ with σ_k SOS only depends on I_k

Uniform H_k involve products $\|\mathbf{x}(I)\|_2^2$ for $I \in \{I_k, \hat{I}_k, \hat{I}_i : s_i = k\}$

More and more applications!

Robust Geometric Perception [Yang & Carlone '20]

More and more applications!

Robust Geometric Perception [Yang & Carlone '20]

Polynomial matrix inequalities [Zheng & Fantuzzi '20]

More and more applications!

Robust Geometric Perception [Yang & Carlone ’20]

Polynomial matrix inequalities [Zheng & Fantuzzi ’20]

Region of attraction [Tacchi et al., Schlosser et al. ’21]

More and more applications!

Robust Geometric Perception [Yang & Carlone '20]

Polynomial matrix inequalities [Zheng & Fantuzzi '20]

Region of attraction [Tacchi et al., Schlosser et al. '21]

Volume computation [Tacchi et al. '21]

More and more applications!

Robust Geometric Perception [Yang & Carlone '20]

Polynomial matrix inequalities [Zheng & Fantuzzi '20]

Region of attraction [Tacchi et al., Schlosser et al. '21]

Volume computation [Tacchi et al. '21]

Robustness of implicit deep networks [Chen et al. '21]

Sparse SDP

Correlative sparsity in POP

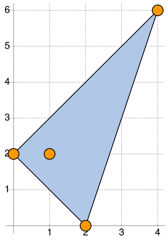
Term sparsity in POP

Conclusion & further topics

Tutorial session

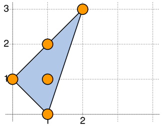
Term sparsity via Newton polytope

$f = 4x_1^4x_2^6 + x_1^2 - x_1x_2^2 + x_2^2$
spt(f) = $\{(4,6),(2,0),(1,2),(0,2)\}$



Newton polytope $\mathscr{B} = \operatorname{conv}(\text{spt}(f))$

Squares in SOS decomposition $\subseteq \frac{\mathscr{B}}{2} \cap \mathbb{N}^n$
[Reznick '78]



$$f = \begin{pmatrix} x_1 & x_2 & x_1x_2 & x_1x_2^2 & x_1^2x_2^3 \end{pmatrix} \underbrace{Q}_{\succcurlyeq 0} \begin{pmatrix} x_1 \\ x_2 \\ x_1x_2 \\ x_1x_2^2 \\ x_1^2x_2^3 \end{pmatrix}$$

Term sparsity: the unconstrained case

$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 + 6x_3^2 + 18x_2^2x_3 - 54x_2x_3^2 + 142x_2^2x_3^2$$

[Reznick '78] $\rightarrow f = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_1x_2 & x_2x_3 \end{pmatrix} \underbrace{Q}_{\succcurlyeq 0} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$

$\rightsquigarrow \frac{6\times 7}{2} = 21$ "unknown" entries in Q

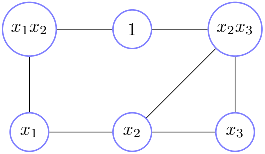
Term sparsity: the unconstrained case

$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 + 6x_3^2 + 18x_2^2x_3 - 54x_2x_3^2 + 142x_2^2x_3^2$$

[Reznick '78] $\to f = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_1x_2 & x_2x_3 \end{pmatrix} \underbrace{Q}_{\succcurlyeq 0} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$

$\leadsto \frac{6\times 7}{2} = 21$ "unknown" entries in Q

💡 **Term** sparsity pattern graph G



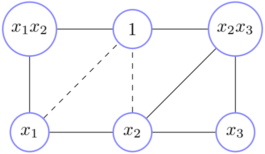
Term sparsity: the unconstrained case

$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 + 6x_3^2 + 18x_2^2x_3 - 54x_2x_3^2 + 142x_2^2x_3^2$$

[Reznick '78] $\to f = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_1x_2 & x_2x_3 \end{pmatrix} \underbrace{Q}_{\succcurlyeq 0} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$

$\rightsquigarrow \frac{6\times 7}{2} = 21$ "unknown" entries in Q

Term sparsity pattern graph G
+ chordal extension G'



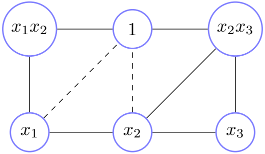
Term sparsity: the unconstrained case

$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 + 6x_3^2 + 18x_2^2x_3 - 54x_2x_3^2 + 142x_2^2x_3^2$$

[Reznick '78] $\to f = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_1x_2 & x_2x_3 \end{pmatrix} \underbrace{Q}_{\succcurlyeq 0} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$

$\rightsquigarrow \frac{6\times 7}{2} = 21$ "unknown" entries in Q

Term sparsity pattern graph G
+ chordal extension G'



Replace Q by $Q_{G'}$ with nonzero entries at edges of G'

$\rightsquigarrow$ 6 + 9 = 15 "unknown" entries in $Q_{G'}$

Term sparsity: the constrained case

At step r of the hierarchy, tsp graph G has

Nodes V = monomials of degree $\leqslant r$

Term sparsity: the constrained case

At step r of the hierarchy, tsp graph G has

Nodes V = monomials of degree $\leqslant r$
Edges E with

$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \operatorname{supp} f \bigcup \operatorname{supp} g_j \bigcup_{|\alpha| \leqslant r} 2\alpha$$

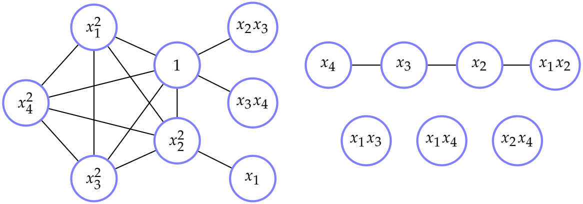
Term sparsity: the constrained case

At step r of the hierarchy, tsp graph G has

Nodes V = monomials of degree $\leqslant r$
Edges E with

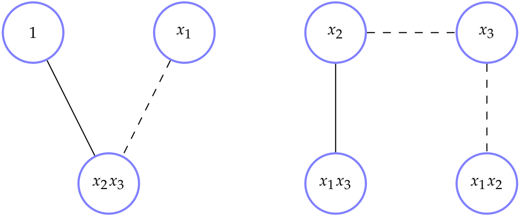
$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \operatorname{supp} f \bigcup \operatorname{supp} g_j \bigcup_{|\alpha| \leqslant r} 2\alpha$$

An example with $r = 2$
$f = x_1^4 + x_1 x_2^2 + x_2 x_3 + x_3^2 x_4^2$
$g_1 = 1 - x_1^2 - x_2^2 - x_3^2 \quad g_2 = 1 - x_3 x_4$



Term sparsity: support extension

$\alpha' + \beta' = \alpha + \beta$ and $(\alpha, \beta) \in E \Rightarrow (\alpha', \beta') \in E$



Term sparsity: the constrained case

At step r of the hierarchy, tsp graph G has

Nodes V = monomials of degree $\leqslant r$

Term sparsity: the constrained case

At step r of the hierarchy, tsp graph G has

Nodes V = monomials of degree $\leqslant r$
Edges E with

$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \operatorname{supp} f \bigcup \operatorname{supp} g_j \bigcup_{|\alpha| \leqslant r} 2\alpha$$

$\rightsquigarrow$ support extension

Term sparsity: the constrained case

At step r of the hierarchy, tsp graph G has

Nodes V = monomials of degree $\leqslant r$
Edges E with

$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \operatorname{supp} f \bigcup \operatorname{supp} g_j \bigcup_{|\alpha| \leqslant r} 2\alpha$$

$\rightsquigarrow$ support extension $\rightsquigarrow$ chordal extension G'

Term sparsity: the constrained case

At step r of the hierarchy, tsp graph G has

Nodes V = monomials of degree $\leqslant r$
Edges E with

$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \operatorname{supp} f \bigcup \operatorname{supp} g_j \bigcup_{|\alpha| \leqslant r} 2\alpha$$

$\rightsquigarrow$ support extension $\rightsquigarrow$ chordal extension G'

By iteratively performing support extension & chordal extension

$$G^{(1)} = G' \subseteq \cdots \subseteq G^{(s)} \subseteq G^{(s+1)} \subseteq \cdots$$

Two-level hierarchy of lower bounds for $f_{\min}$, indexed by sparse order s and relaxation order r

Term sparsity: primal moment relaxations

Let G' be a chordal extension of G with maximal cliques (C_i)

$$C_i \longmapsto \mathbf{M}_{C_i}(\mathbf{y})$$

Term sparsity: primal moment relaxations

Let G' be a chordal extension of G with maximal cliques (C_i)

$$C_i \longmapsto \mathbf{M}_{C_i}(\mathbf{y})$$

In the moment relaxation,

$$\mathbf{M}_r(\mathbf{y}) \succcurlyeq 0 \longrightarrow \mathbf{M}_{C_i}(\mathbf{y}) \succcurlyeq 0$$

Term sparsity: primal moment relaxations

Let G' be a chordal extension of G with maximal cliques (C_i)

$$C_i \longmapsto \mathbf{M}_{C_i}(\mathbf{y})$$

In the moment relaxation,

$$\mathbf{M}_r(\mathbf{y}) \succcurlyeq 0 \longrightarrow \mathbf{M}_{C_i}(\mathbf{y}) \succcurlyeq 0$$

Similarly for the localizing matrices $\mathbf{M}_{r-r_j}(g_j\,\mathbf{y})$

Term sparsity: primal moment relaxations

Let G' be a chordal extension of G with maximal cliques (C_i)

$$C_i \longmapsto \mathbf{M}_{C_i}(\mathbf{y})$$

In the moment relaxation,

$$\mathbf{M}_r(\mathbf{y}) \succcurlyeq 0 \longrightarrow \mathbf{M}_{C_i}(\mathbf{y}) \succcurlyeq 0$$

Similarly for the localizing matrices $\mathbf{M}_{r-r_j}(g_j\,\mathbf{y})$

Each constraint $G_j \rightsquigarrow G_j^{(s)} \rightsquigarrow$ cliques $C_{j,i}^{(s)}$

Term sparsity: primal moment relaxations

Let $C_{j,i}^{(s)}$ be the maximal cliques of $G_j^{(s)}$. For each $s \geq 1$

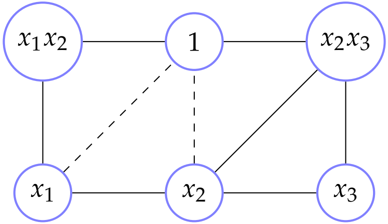
$$
\begin{aligned}
f_{\text{ts}}^{r,s} = \quad \inf \quad & \textstyle\sum_\alpha f_\alpha y_\alpha \\
\text{s.t.} \quad & \mathbf{M}_{C_{0,i}^{(s)}}(\mathbf{y}) \succcurlyeq 0 \\
& \mathbf{M}_{C_{j,i}^{(s)}}(g_j\,\mathbf{y}) \succcurlyeq 0 \\
& y_0 = 1
\end{aligned}
$$

💡 dual yields the TSSOS hierarchy

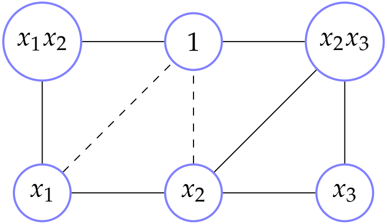
A two-level hierarchy of lower bounds

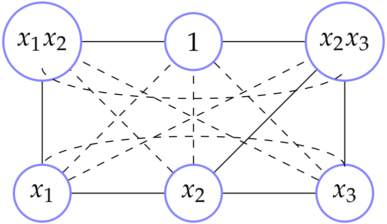
$$
\begin{array}{ccccccc}
f_{\mathrm{ts}}^{r_{\min},1} & \leq & f_{\mathrm{ts}}^{r_{\min},2} & \leq & \cdots & \leq & f^{r_{\min}} \\
\wedge| & & \wedge| & & & & \wedge| \\
f_{\mathrm{ts}}^{r_{\min}+1,1} & \leq & f_{\mathrm{ts}}^{r_{\min}+1,2} & \leq & \cdots & \leq & f^{r_{\min}+1} \\
\wedge| & & \wedge| & & & & \wedge| \\
\vdots & & \vdots & & \vdots & & \vdots \\
\wedge| & & \wedge| & & & & \wedge| \\
f_{\mathrm{ts}}^{r,1} & \leq & f_{\mathrm{ts}}^{r,2} & \leq & \cdots & \leq & f^{r} \\
\wedge| & & \wedge| & & & & \wedge| \\
\vdots & & \vdots & & \vdots & & \vdots
\end{array}
$$

Different choices of chordal extensions



Different choices of chordal extensions





Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

Fixing a sparse order s, the sequence $(f_{\text{ts}}^{r,s})_{r \geq r_{\min}}$ is monotonically nondecreasing.

Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

Fixing a sparse order s, the sequence $(f_{\text{ts}}^{r,s})_{r \geq r_{\min}}$ is monotonically nondecreasing.

Fixing a relaxation order r, the sequence $(f_{\text{ts}}^{r,s})_{s \geq 1} \uparrow f^r$ in finitely many steps if the maximal chordal extension is used.

Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

Fixing a sparse order s, the sequence $(f_{\text{ts}}^{r,s})_{r \geq r_{\min}}$ is monotonically nondecreasing.

💡 Fixing a relaxation order r, the sequence $(f_{\text{ts}}^{r,s})_{s \geq 1} \uparrow f^r$ in finitely many steps if the maximal chordal extension is used.

💡 The block structures converge to the one determined by the sign symmetries if the maximal chordal extension is used.

Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

Fixing a sparse order s, the sequence $(f_{\text{ts}}^{r,s})_{r \geq r_{\min}}$ is monotonically nondecreasing.

Fixing a relaxation order r, the sequence $(f_{\text{ts}}^{r,s})_{s \geq 1} \uparrow f^r$ in finitely many steps if the maximal chordal extension is used.

The block structures converge to the one determined by the sign symmetries if the maximal chordal extension is used.

$f = 1 + x_1^2 x_2^4 + x_1^4 x_2^2 + x_1^4 x_2^4 - x_1 x_2^2 - 3 x_1^2 x_2^2$
Newton polytope $\rightsquigarrow \mathscr{B} = (1 \quad x_1 x_2 \quad x_1 x_2^2 \quad x_1^2 x_2 \quad x_1^2 x_2^2)$

Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

Fixing a sparse order s, the sequence $(f_{\text{ts}}^{r,s})_{r \geq r_{\min}}$ is monotonically nondecreasing.

Fixing a relaxation order r, the sequence $(f_{\text{ts}}^{r,s})_{s \geq 1} \uparrow f^r$ in finitely many steps if the maximal chordal extension is used.

The block structures converge to the one determined by the sign symmetries if the maximal chordal extension is used.

$f = 1 + x_1^2 x_2^4 + x_1^4 x_2^2 + x_1^4 x_2^4 - x_1 x_2^2 - 3 x_1^2 x_2^2$

Newton polytope $\rightsquigarrow \mathscr{B} = (1 \quad x_1 x_2 \quad x_1 x_2^2 \quad x_1^2 x_2 \quad x_1^2 x_2^2)$

$x_2 \mapsto -x_2$

Sign-symmetries blocks	$(1 \quad x_1 x_2^2 \quad x_1^2 x_2^2)$	$(x_1 x_2 \quad x_1^2 x_2)$	
TSSOS blocks	$(1 \quad x_1 x_2^2 \quad x_1^2 x_2^2)$	$(x_1 x_2)$	$(x_1^2 x_2)$

A second key message

💡 **TSSOS preserves the block structure related to sign-symmetries** 💡

Comparison with (S)DSOS

Let f be a nonnegative polynomial of degree $2d$
f is SOS $\Leftrightarrow f = \mathbf{v}^T \mathbf{Q} \mathbf{v}$ with $\mathbf{Q} \succcurlyeq 0 \rightsquigarrow$ semidefinite program
where $\mathbf{v}$ contains $1, x_1, \dots, x_n, x_1^2, \dots, x_n^d$

Comparison with (S)DSOS

Let f be a nonnegative polynomial of degree $2d$
f is SOS $\Leftrightarrow f = \mathbf{v}^T \mathbf{Q} \mathbf{v}$ with $\mathbf{Q} \succcurlyeq 0 \rightsquigarrow$ semidefinite program
where $\mathbf{v}$ contains $1, x_1, \ldots, x_n, x_1^2, \ldots, x_n^d$

To reduce the number of "unknown" entries in $\mathbf{Q}$, one can force:
[Ahmadi & Majumdar '14]

1. $\mathbf{Q}$ diagonally dominant: $\mathbf{Q}_{ii} \geqslant \sum_{j \neq i} \mathbf{Q}_{ij} \rightsquigarrow$ linear program
2. $\mathbf{Q}$ $\sim$ to a diag. dominant matrix $\rightsquigarrow$ second-order program

Comparison with (S)DSOS

Let f be a nonnegative polynomial of degree $2d$
f is SOS $\Leftrightarrow f = \mathbf{v}^T \mathbf{Q} \mathbf{v}$ with $\mathbf{Q} \succcurlyeq 0 \rightsquigarrow$ semidefinite program
where $\mathbf{v}$ contains $1, x_1, \dots, x_n, x_1^2, \dots, x_n^d$

To reduce the number of "unknown" entries in $\mathbf{Q}$, one can force: [Ahmadi & Majumdar '14]

1. $\mathbf{Q}$ diagonally dominant: $\mathbf{Q}_{ii} \geqslant \sum_{j \neq i} \mathbf{Q}_{ij} \rightsquigarrow$ linear program
2. $\mathbf{Q}$ $\sim$ to a diag. dominant matrix $\rightsquigarrow$ second-order program

Theorem [Lasserre Magron Wang '21]

The first TSSOS relaxation is always more accurate than the SDSOS relaxation

Combining correlative & term sparsity

1. Partition the variables w.r.t. the maximal cliques of the csp graph

Combining correlative & term sparsity

1. Partition the variables w.r.t. the maximal cliques of the csp graph
2. For each subsystem involving variables from one maximal clique, apply TSSOS

Combining correlative & term sparsity

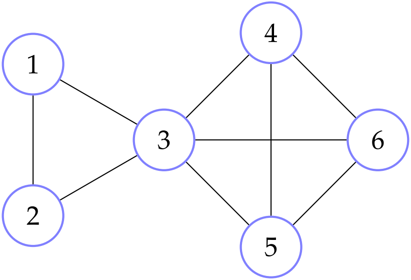
1. Partition the variables w.r.t. the maximal cliques of the csp graph
2. For each subsystem involving variables from one maximal clique, apply TSSOS

💡 a two-level CS-TSSOS hierarchy of lower bounds for $f_{\min}$

Combining correlative & term sparsity

$$f = 1 + \sum_{i=1}^{6} x_i^4 + x_1x_2x_3 + x_3x_4x_5 + x_3x_4x_6 + x_3x_5x_6 + x_4x_5x_6$$

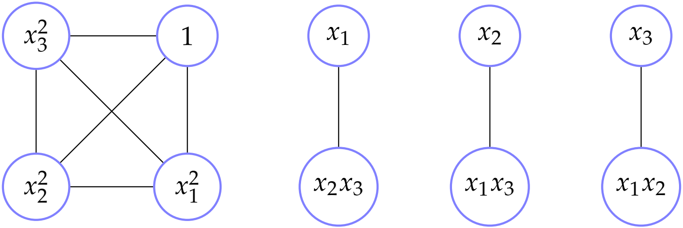
csp graph



Combining correlative & term sparsity

$$f = 1 + \sum_{i=1}^{6} x_i^4 + x_1x_2x_3 + x_3x_4x_5 + x_3x_4x_6 + x_3x_5x_6 + x_4x_5x_6$$

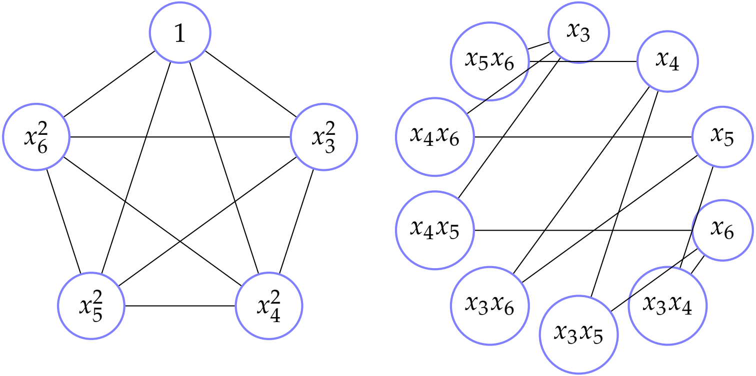
tsp graph for the first clique



Combining correlative & term sparsity

$$f = 1 + \sum_{i=1}^{6} x_i^4 + x_1x_2x_3 + x_3x_4x_5 + x_3x_4x_6 + x_3x_5x_6 + x_4x_5x_6$$

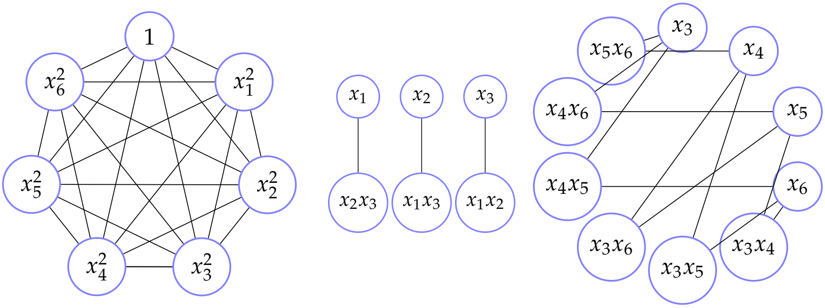
tsp graph for the second clique



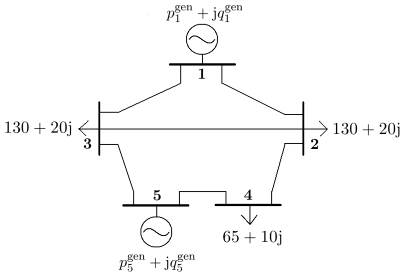
Combining correlative & term sparsity

$$f = 1 + \sum_{i=1}^{6} x_i^4 + x_1x_2x_3 + x_3x_4x_5 + x_3x_4x_6 + x_3x_5x_6 + x_4x_5x_6$$

tsp graph without correlative sparsity



Application to optimal power-flow



Optimal Powerflow $n \simeq \mathbf{10^3}$

[Josz et al. '18]

$$
\begin{cases}
\inf\limits_{V_i, S_s^g, S_{ij}} & \sum_{s \in G} (\mathbf{c}_{2s} (\Re(S_s^g))^2 + \mathbf{c}_{1s} \Re(S_s^g) + \mathbf{c}_{0s}) \\
\text{s.t.} & \angle V_{\text{ref}} = 0, \\
& \mathbf{S}_s^{gl} \leq S_s^g \leq \mathbf{S}_s^{gu} \ \forall s \in G, \quad \boldsymbol{v}_i^l \leq |V_i| \leq \boldsymbol{v}_i^u \ \forall i \in N \\
& \sum_{s \in G_i} S_s^g - \mathbf{S}_i^d - \mathbf{Y}_i^s |V_i|^2 = \sum_{(i,j) \in E_i \cup E_i^R} S_{ij}, \quad \forall i \in N \\
& S_{ij} = (\mathbf{Y}_{ij}^* - \mathbf{i} \frac{\mathbf{b}_{ij}^c}{2}) \frac{|V_i|^2}{|\mathbf{T}_{ij}|^2} - \mathbf{Y}_{ij}^* \frac{V_i V_j^*}{\mathbf{T}_{ij}}, \quad S_{ji} = \cdots \\
& |S_{ij}| \leq \mathbf{s}_{ij}^u, \ \boldsymbol{\theta}_{ij}^{\Delta l} \leq \angle(V_i V_j^*) \leq \boldsymbol{\theta}_{ij}^{\Delta u}, \quad \forall (i,j) \in E
\end{cases}
$$

Application to optimal power-flow

mb = the maximal size of blocks

m = number of constraints

n	m	CS ($r = 2$)			CS+TS ($r = 2, s = 1$)		
		mb	time (s)	gap	mb	time (s)	gap
114	315	66	5.59	0.39%	31	2.01	0.73%
348	1809	253	–	–	34	278	0.05%
766	3322	153	585	0.68%	44	33.9	0.77%
1112	4613	496	–	–	31	410	0.25%
4356	18257	378	–	–	27	934	0.51%
6698	29283	1326	–	–	76	1886	0.47%

Application to networked systems stability

Lyapunov function

$$f = \sum_{i=1}^{N} a_i (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} b_{ik} x_i^2 x_k^2 \quad a_i \in [1, 2] \quad b_{ik} \in [\frac{0.5}{N}, \frac{1.5}{N}]$$

$\rightsquigarrow \binom{N+2}{2}(\binom{N+2}{2} + 1)/2$ "unknown" entries in $Q = 231$ for $N = 5$

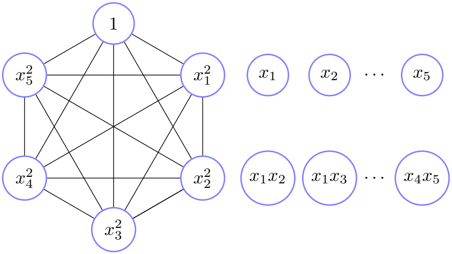
Application to networked systems stability

Lyapunov function

$$f = \sum_{i=1}^{N} a_i (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} b_{ik} x_i^2 x_k^2 \quad a_i \in [1, 2] \quad b_{ik} \in [\frac{0.5}{N}, \frac{1.5}{N}]$$

$\rightsquigarrow \binom{N+2}{2}(\binom{N+2}{2} + 1)/2$ “unknown” entries in $Q = 231$ for $N = 5$

tsp graph G



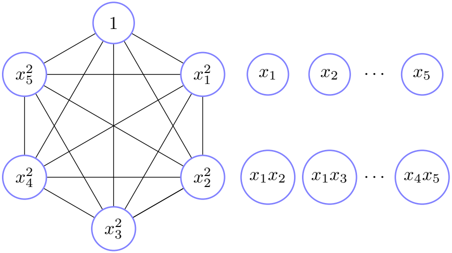
Application to networked systems stability

Lyapunov function

$$f = \sum_{i=1}^{N} a_i (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} b_{ik} x_i^2 x_k^2 \quad a_i \in [1,2] \quad b_{ik} \in [\frac{0.5}{N}, \frac{1.5}{N}]$$

$\rightsquigarrow \binom{N+2}{2}(\binom{N+2}{2}+1)/2$ "unknown" entries in $Q = 231$ for $N = 5$

💡 **tsp** graph G



$\rightsquigarrow (N+1)^2$ "unknown" entries in $Q_G = 36$ for $N = 5$

Proof that $f \geqslant 0$ for $N = 80$ in ~ 10 seconds!

Application to networked systems stability

Duffing oscillator Hamiltonian $V = \sum_{i=1}^{N} a_i(\frac{x_i^2}{2} - \frac{x_i^4}{4}) + \frac{1}{8}\sum_{i,k=1}^{N} b_{ik}(x_i - x_k)^4$

On which domain $V > 0$?

$$f = V - \sum_{i=1}^{N} \underbrace{\lambda_i}_{>0} x_i^2(g - x_i^2) \geqslant 0$$

$\implies V > 0$ when $x_i^2 < g$

Application to networked systems stability

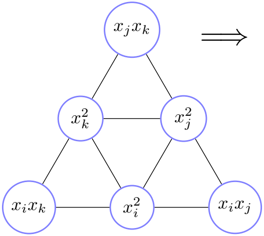
Duffing oscillator Hamiltonian $V = \sum_{i=1}^{N} a_i(\frac{x_i^2}{2} - \frac{x_i^4}{4}) + \frac{1}{8}\sum_{i,k=1}^{N} b_{ik}(x_i - x_k)^4$

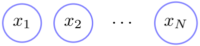
On which domain $V > 0$?

$$f = V - \sum_{i=1}^{N} \underbrace{\lambda_i}_{>0} x_i^2(g - x_i^2) \geqslant 0$$

$$\implies V > 0 \text{ when } x_i^2 < g$$

tsp graph G





Application to networked systems stability

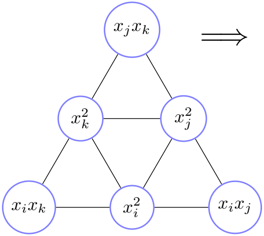
Duffing oscillator Hamiltonian $V = \sum_{i=1}^{N} a_i \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \frac{1}{8} \sum_{i,k=1}^{N} b_{ik} (x_i - x_k)^4$

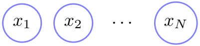
On which domain $V > 0$?

$$f = V - \sum_{i=1}^{N} \underbrace{\lambda_i}_{>0} x_i^2 (g - x_i^2) \geqslant 0$$

$\implies V > 0$ when $x_i^2 < g$

tsp graph G





$\rightsquigarrow \frac{N(N+1)}{2} + 6\binom{N}{2} + N$ "unknown" entries in $Q_G = 80$ for $N = 5$

Proof that $f \geqslant 0$ for $N = 50$ in ~ 1 second!

Application to joint spectral radius (JSR)

Given $\mathcal{A} = \{A_1, \ldots, A_m\} \subseteq \mathbb{R}^{n \times n}$, the JSR is

$$\rho(\mathcal{A}) := \lim_{k \to \infty} \max_{\sigma \in \{1,\ldots,m\}^k} ||A_{\sigma_1} A_{\sigma_2} \cdots A_{\sigma_k}||^{\frac{1}{k}}$$

Application to joint spectral radius (JSR)

Given $\mathcal{A} = \{A_1, \ldots, A_m\} \subseteq \mathbb{R}^{n \times n}$, the JSR is

$$\rho(\mathcal{A}) := \lim_{k \to \infty} \max_{\sigma \in \{1,\ldots,m\}^k} ||A_{\sigma_1} A_{\sigma_2} \cdots A_{\sigma_k}||^{\frac{1}{k}}$$

Tons of applications:

- stability of switched linear systems
- continuity of wavelet functions
- trackability of graphs
- $\cdots$

Application to joint spectral radius (JSR)

Given $\mathcal{A} = \{A_1, \ldots, A_m\} \subseteq \mathbb{R}^{n \times n}$, the JSR is

$$\rho(\mathcal{A}) := \lim_{k \to \infty} \max_{\sigma \in \{1,\ldots,m\}^k} ||A_{\sigma_1} A_{\sigma_2} \cdots A_{\sigma_k}||^{\frac{1}{k}}$$

Tons of applications:

- stability of switched linear systems
- continuity of wavelet functions
- trackability of graphs
- $\cdots$

. . . NP-hard to compute/approximate

Application to joint spectral radius (JSR)

Theorem [Parrilo & Jadbabaie '08]

Given $\mathcal{A} = \{A_1, \ldots, A_m\} \subseteq \mathbb{R}^{n \times n}$, if a positive definite form f of degree $2r$ satisfies

$$f(A_i \mathbf{x}) \leqslant \gamma^{2r} f(\mathbf{x}) \quad \forall i, \mathbf{x}$$

Then $\rho(\mathcal{A}) \leqslant \gamma$

Application to joint spectral radius (JSR)

Theorem [Parrilo & Jadbabaie '08]

Given $\mathcal{A} = \{A_1, \ldots, A_m\} \subseteq \mathbb{R}^{n \times n}$, if a positive definite form f of degree $2r$ satisfies

$$f(A_i \mathbf{x}) \leqslant \gamma^{2r} f(\mathbf{x}) \quad \forall i, \mathbf{x}$$

Then $\rho(\mathcal{A}) \leqslant \gamma$

$$\rho(\mathcal{A}) \leqslant \rho^r(\mathcal{A}) = \inf_{f \in \mathbb{R}[\mathbf{x}]_{2r}, \gamma} \gamma$$

$$\text{s.t.} \begin{cases} f(\mathbf{x}) - ||\mathbf{x}||_2^{2r} \text{ SOS} \\ \gamma^{2r} f(\mathbf{x}) - f(A_i \mathbf{x}) \text{ SOS} \end{cases}$$

Application to joint spectral radius (JSR)

Theorem [Parrilo & Jadbabaie '08]

Given $\mathcal{A} = \{A_1, \ldots, A_m\} \subseteq \mathbb{R}^{n \times n}$, if a positive definite form f of degree $2r$ satisfies

$$f(A_i \mathbf{x}) \leqslant \gamma^{2r} f(\mathbf{x}) \quad \forall i, \mathbf{x}$$

Then $\rho(\mathcal{A}) \leqslant \gamma$

$$\rho(\mathcal{A}) \leqslant \rho^r(\mathcal{A}) = \inf_{f \in \mathbb{R}[\mathbf{x}]_{2r}, \gamma} \gamma$$

$$\text{s.t.} \begin{cases} f(\mathbf{x}) - ||\mathbf{x}||_2^{2r} \text{ SOS} \\ \gamma^{2r} f(\mathbf{x}) - f(A_i \mathbf{x}) \text{ SOS} \end{cases}$$

Bisection on γ + SDP

Application to joint spectral radius (JSR)

- At fixed r, replace f by a “term sparse” f

Application to joint spectral radius (JSR)

At fixed r, replace f by a "term sparse" f

$f_0 = \sum_{j=1}^{n} c_j x_j^{2r}$ with support $\mathscr{A}^{(0)}$

Application to joint spectral radius (JSR)

At fixed r, replace f by a "term sparse" f

$f_0 = \sum_{j=1}^{n} c_j x_j^{2r}$ with support $\mathscr{A}^{(0)}$
Recursively, $f_{s-1} = \sum_{\alpha \in \mathscr{A}^{(s-1)}} c_\alpha \mathbf{x}^\alpha$ and

$$\mathscr{A}^{(s)} = \mathscr{A}^{(s-1)} \cup \bigcup_i \operatorname{supp}(f_{s-1}(A_i \mathbf{x})) \quad \mathscr{A}_i^{(s)} = \mathscr{A}^{(s)} \cup \operatorname{supp}(f_s(A_i \mathbf{x}))$$

Application to joint spectral radius (JSR)

At fixed r, replace f by a "term sparse" f

$f_0 = \sum_{j=1}^{n} c_j x_j^{2r}$ with support $\mathscr{A}^{(0)}$
Recursively, $f_{s-1} = \sum_{\alpha \in \mathscr{A}^{(s-1)}} c_\alpha \mathbf{x}^\alpha$ and

$$\mathscr{A}^{(s)} = \mathscr{A}^{(s-1)} \cup \bigcup_i \mathrm{supp}(f_{s-1}(A_i \mathbf{x})) \quad \mathscr{A}_i^{(s)} = \mathscr{A}^{(s)} \cup \mathrm{supp}(f_s(A_i \mathbf{x}))$$

Theorem: Sparse JSR [Maggio Magron Wang '21]

$$\rho(\mathcal{A}) \leqslant \rho^r(\mathcal{A}) \leqslant \rho^{r,s}(\mathcal{A}) = \inf_{f \in \mathbb{R}[\mathscr{A}^{(s)}], \gamma} \gamma$$

$$\text{s.t. } \begin{cases} f(\mathbf{x}) - ||\mathbf{x}||_2^{2r} \text{ SOS}(\mathscr{A}^{(s)}) \\ \gamma^{2r} f(\mathbf{x}) - f(A_i \mathbf{x}) \text{ SOS}(\mathscr{A}_i^{(s)}) \end{cases}$$

Application to joint spectral radius (JSR)

Closed-loop system evolves according to either a completed or a missed computation (A_H or A_M): $\mathcal{A} = \{A_H {A_M}^i \mid i < m\}$

Application to joint spectral radius (JSR)

Closed-loop system evolves according to either a completed or a missed computation (A_H or A_M): $\mathcal{A} = \{A_H {A_M}^i \mid i < m\}$

System asymptotically stable $\Leftrightarrow \rho(\mathcal{A}) < 1$

Application to joint spectral radius (JSR)

Closed-loop system evolves according to either a completed or a missed computation (A_H or A_M): $\mathcal{A} = \{A_H {A_M}^i \mid i < m\}$

System asymptotically stable $\Leftrightarrow \rho(\mathcal{A}) < 1$

(un)stability test with 10 matrices & $n = 25$ or 2 matrices & $n = 100$ intractable with the dense JSR

Application to joint spectral radius (JSR)

Closed-loop system evolves according to either a completed or a missed computation (A_H or A_M): $\mathcal{A} = \{A_H {A_M}^i \mid i < m\}$

System asymptotically stable $\Leftrightarrow \rho(\mathcal{A}) < 1$

(un)stability test with 10 matrices & $n = 25$ or 2 matrices & $n = 100$ intractable with the dense JSR

takes less than 10 seconds with the Sparse JSR!

Conclusion

SPARSITY EXPLOITING CONVERGING HIERARCHIES to minimize polynomials, eigenvalue/trace, JSR

Conclusion

SPARSITY EXPLOITING CONVERGING HIERARCHIES to minimize polynomials, eigenvalue/trace, JSR

FAST IMPLEMENTATION IN JULIA: TSSOS, NCTSSOS, SparseJSR

Conclusion

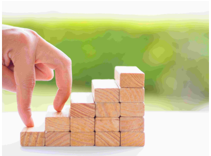
SPARSITY EXPLOITING CONVERGING HIERARCHIES to minimize polynomials, eigenvalue/trace, JSR

FAST IMPLEMENTATION IN JULIA: TSSOS, NCTSSOS, SparseJSR

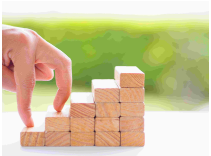
💡 Combine correlative & term sparsity for problems with $n = 10^3$

Further topics

Convergence rate of SPARSE hierarchies?



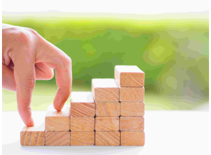
Further topics



Convergence rate of SPARSE hierarchies?

💡 (smart) solution extraction for term sparsity

Further topics

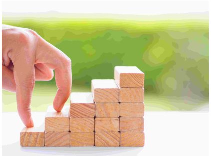


Convergence rate of SPARSE hierarchies?

💡 (smart) solution extraction for term sparsity

Numerical conditioning of sparse SDP?

Further topics



Convergence rate of SPARSE hierarchies?

💡 (smart) solution extraction for term sparsity

Numerical conditioning of sparse SDP?

💡 Tons of applications . . .

Further topics: deep learning

💡 "Direct" certification of a classifier with 1 hidden layer

$$\max_{\mathbf{x},\mathbf{z}} \quad (\mathbf{C}^{i,:} - \mathbf{C}^{k,:})\mathbf{z}$$

$$\text{s.t.} \quad \begin{cases} \mathbf{z} = \mathrm{ReLU}(\mathbf{A}\mathbf{x} + \mathbf{b}) \\ ||\mathbf{x} - \mathbf{x}_0|| \leq \epsilon \end{cases}$$

Further topics: deep learning

- "Direct" certification of a classifier with 1 hidden layer

$$\max_{\mathbf{x},\mathbf{z}} \quad (\mathbf{C}^{i,:} - \mathbf{C}^{k,:})\mathbf{z}$$

$$\text{s.t.} \quad \begin{cases} \mathbf{z} = \mathrm{ReLU}(\mathbf{A}\mathbf{x} + \mathbf{b}) \\ ||\mathbf{x} - \mathbf{x}_0|| \leq \epsilon \end{cases}$$

- Monotone equilibrium networks [Winston Kolter '20]

$$\mathbf{z} = \mathrm{ReLU}(\mathbf{A}\mathbf{x} + \mathbf{b}) \rightarrow \mathbf{z} = \mathrm{ReLU}(\mathbf{W}\mathbf{z} + \mathbf{A}\mathbf{x} + \mathbf{b})$$

Further topics: deep learning

- “Direct” certification of a classifier with 1 hidden layer

$$\max_{\mathbf{x},\mathbf{z}} \quad (\mathbf{C}^{i,:} - \mathbf{C}^{k,:})\mathbf{z}$$
$$\text{s.t.} \quad \begin{cases} \mathbf{z} = \mathrm{ReLU}(\mathbf{A}\mathbf{x} + \mathbf{b}) \\ ||\mathbf{x} - \mathbf{x}_0|| \leq \epsilon \end{cases}$$

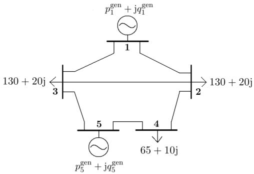
- Monotone equilibrium networks [Winston Kolter ’20]

$$\mathbf{z} = \mathrm{ReLU}(\mathbf{A}\mathbf{x} + \mathbf{b}) \to \mathbf{z} = \mathrm{ReLU}(\mathbf{W}\mathbf{z} + \mathbf{A}\mathbf{x} + \mathbf{b})$$

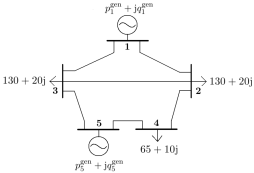
- “Indirect” with Lipschitz constant/ellipsoid approximation

Further topics: power systems

Solving Alternative Current OPF to **global optimality**
$\rightarrow$ benchmarks [PGLIB '18] with up to **25 000 buses**!



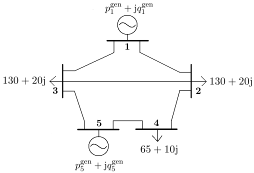
Further topics: power systems



Solving Alternative Current OPF to **global optimality**
$\rightarrow$ benchmarks [PGLIB '18] with up to **25 000 buses**!

COMPLEX vs REAL hierarchy of relaxations?
[D'Angelo Putinar '09, Josz et al. '18, Magron Wang '21]

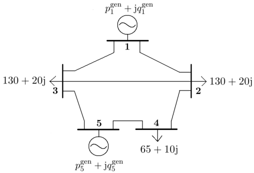
Further topics: power systems



Solving Alternative Current OPF to **global optimality**
$\rightarrow$ benchmarks [PGLIB '18] with up to **25 000 buses**!

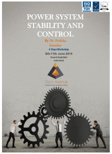
COMPLEX vs REAL hierarchy of relaxations?
[D'Angelo Putinar '09, Josz et al. '18, Magron Wang '21]
6515_RTE $\rightarrow n = 7000$ complex variables (14000 real variables)
solved at 0.6% gap within 3 hours on PFCALCUL at LAAS

Further topics: power systems



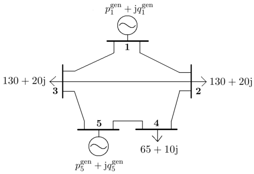
Solving Alternative Current OPF to **global optimality**
$\rightarrow$ benchmarks [PGLIB '18] with up to **25 000 buses**!

COMPLEX vs REAL hierarchy of relaxations?
[D'Angelo Putinar '09, Josz et al. '18, Magron Wang '21]
6515_RTE $\rightarrow n = 7000$ complex variables (14000 real variables)
solved at 0.6% gap within 3 hours on PFCALCUL at LAAS



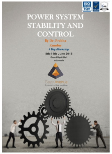
STABILITY OF LARGE-SCALE POWER SYSTEMS $\rightarrow$ reachability analysis of continuous-time systems **Sparse** [Kundur '07]

Further topics: power systems



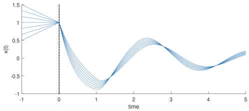
Solving Alternative Current OPF to **global optimality**
$\to$ benchmarks [PGLIB '18] with up to **25 000 buses**!

COMPLEX vs REAL hierarchy of relaxations?
[D'Angelo Putinar '09, Josz et al. '18, Magron Wang '21]
6515_RTE $\to n = 7000$ complex variables (14000 real variables)
solved at 0.6% gap within 3 hours on PFCALCUL at LAAS



STABILITY OF LARGE-SCALE POWER SYSTEMS $\to$ reachability analysis of continuous-time systems **Sparse** [Kundur '07]

TIME DELAY SYSTEMS $\to$ deteriorate controllers of networked power systems **occupation measures**



Thank you for your attention!

https://homepages.laas.fr/vmagron

GITHUB:TSSOS

References

- Gavril. Algorithms for minimum coloring, maximum clique, minimum covering by cliques, and maximum independent set of a chordal graph. SIAM Comp., 1972
- Griewank & Toint. Numerical experiments with partially separable optimization problems. Numerical analysis, 1984
- Agler, Helton, McCullough & Rodman. Positive semidefinite matrices with a given sparsity pattern. Linear algebra & its applications, 1988
- Blair & Peyton. An introduction to chordal graphs and clique trees. Graph theory & sparse matrix computation, 1993
- Vandenberghe & Andersen. Chordal graphs and semidefinite optimization. Foundations & Trends in Optim., 2015

References

- Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity. SIAM Optim., 2006
- Waki, Kim, Kojima & Muramatsu. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. SIAM Optim., 2006
- Magron, Constantinides, & Donaldson. Certified Roundoff Error Bounds Using Semidefinite Programming. Trans. Math. Softw., 2017
- Magron. Interval Enclosures of Upper Bounds of Roundoff Errors Using Semidefinite Programming. Trans. Math. Softw., 2018
- Josz & Molzahn. Lasserre hierarchy for large scale polynomial optimization in real and complex variables. SIAM Optim., 2018
- Weisser, Lasserre & Toh. Sparse-BSOS: a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity. Math. Program., 2018
- Chen, Lasserre, Magron & Pauwels. A sublevel moment-sos hierarchy for polynomial optimization, arxiv:2101.05167

References

- Chen, Lasserre, Magron & Pauwels. Semialgebraic Optimization for Bounding Lipschitz Constants of ReLU Networks. NIPS, 2020
- Chen, Lasserre, Magron & Pauwels. Semialgebraic Representation of Monotone Deep Equilibrium Models and Applications to Certification. arxiv:2106.01453
- Mai, Lasserre & Magron. A sparse version of Reznick's Positivstellensatz. arxiv:2002.05101
- Tacchi, Weisser, Lasserre & Henrion. Exploiting sparsity for semi-algebraic set volume computation. Foundations of Comp. Math., 2021
- Tacchi, Cardozo, Henrion & Lasserre. Approximating regions of attraction of a sparse polynomial differential system. IFAC, 2020
- Schlosser & Korda. Sparse moment-sum-of-squares relaxations for nonlinear dynamical systems with guaranteed convergence. arxiv:2012.05572
- Zheng & Fantuzzi. Sum-of-squares chordal decomposition of polynomial matrix inequalities. arxiv:2007.11410

References

- Klep, Magron & Povh. Sparse Noncommutative Polynomial Optimization. *Math Prog. A*, arxiv:1909.00569 NCSOStools
- Reznick. Extremal PSD forms with few terms. Duke mathematical journal, 1978
- Wang, Magron & Lasserre. TSSOS: A Moment-SOS hierarchy that exploits term sparsity. SIAM Optim., 2021
- Wang, Magron & Lasserre. Chordal-TSSOS: a moment-SOS hierarchy that exploits term sparsity with chordal extension. SIAM Optim., 2021
- Wang, Magron, Lasserre & Mai. CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization. arxiv:2005.02828
- Magron & Wang. TSSOS: a Julia library to exploit sparsity for large-scale polynomial optimization, MEGA, 2021
- Parrilo & Jadbabaie. Approximation of the joint spectral radius using sum of squares. Linear Algebra & its Applications, 2008
- Wang, Maggio & Magron. SparseJSR: A fast algorithm to compute joint spectral radius via sparse sos decompositions. ACC 2021

References

- Vreman, Pazzaglia, Wang, Magron & Maggio. Stability of control systems under extended weakly-hard constraints. arxiv:2101.11312
- Wang & Magron. Exploiting Sparsity in Complex Polynomial Optimization. arxiv:2103.12444
- Wang & Magron. Exploiting term sparsity in Noncommutative Polynomial Optimization. *Computational Optimization & Applications*, arxiv:2010.06956 NCTSSOS
- Navascués, Pironio & Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. New Journal of Physics, 2008
- Klep, Magron & Volčič. Optimization over trace polynomials. Annales Henri Poincaré, 2021

Sparse SDP

Correlative sparsity in POP

Term sparsity in POP

Conclusion & further topics

Tutorial session

Newton polytope

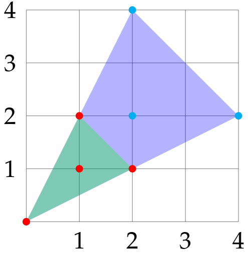
Motzkin $f = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$

1. Compute the Newton polytope of f
2. Show that f is not SOS

Newton polytope

Motzkin $f = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$

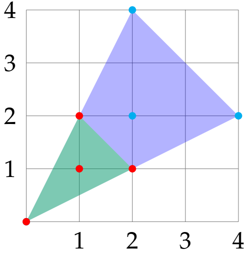
1. Compute the Newton polytope of f
2. Show that f is not SOS



Newton polytope

Motzkin $f = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$

1. Compute the Newton polytope of f
2. Show that f is not SOS

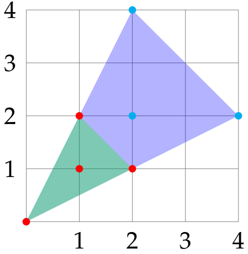


If f SOS then $f = \sum (a x_1^2 x_2 + b x_1 x_2^2 + c x_1 x_2 + d)^2$

Newton polytope

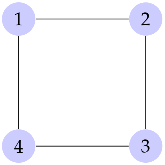
Motzkin $f = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$

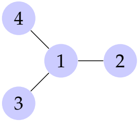
1. Compute the Newton polytope of f
2. Show that f is not SOS

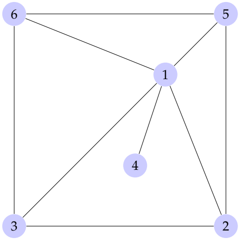


If f SOS then $f = \sum(a x_1^2 x_2 + b x_1 x_2^2 + c x_1 x_2 + d)^2$
never yields $-3x_1^2 x_2^2$

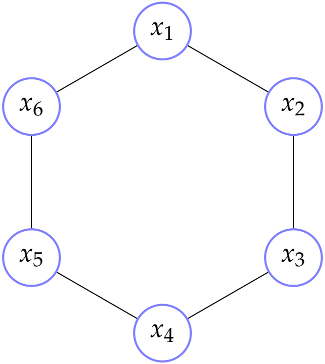
Chordal or not chordal?



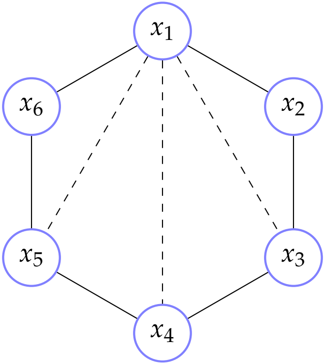




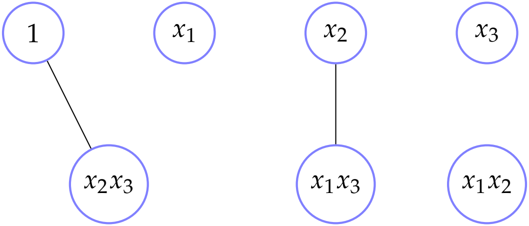
Chordal extension



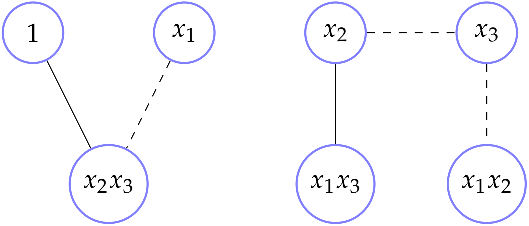
Chordal extension



Support extension



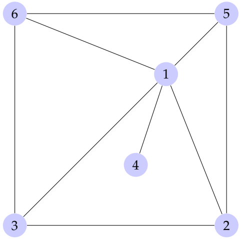
Support extension



How big is CS?

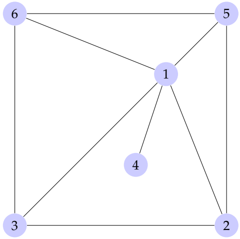
$$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$$

Chordal graph after adding edge $(3, 5)$



How big is CS?

$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$



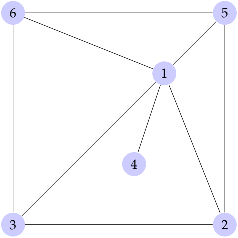
Chordal graph after adding edge $(3, 5)$

How many SDP variables in the dense and sparse relaxation at order $r = 1, 2, 3$?

How big is CS?

$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$

Chordal graph after adding edge $(3, 5)$



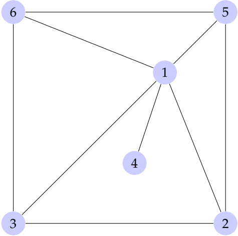
How many SDP variables in the dense and sparse relaxation at order $r = 1, 2, 3$?

maximal cliques $I_1 = \{1, 4\}$ $I_2 = \{1, 2, 3, 5\}$ $I_3 = \{1, 3, 5, 6\}$

How big is CS?

$f(\mathbf{x}) = x_2x_5 + x_3x_6 - x_2x_3 - x_5x_6 + x_1(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6)$

Chordal graph after adding edge $(3,5)$



How many SDP variables in the dense and sparse relaxation at order $r = 1, 2, 3$?

maximal cliques $I_1 = \{1,4\}$ $I_2 = \{1,2,3,5\}$ $I_3 = \{1,3,5,6\}$

$\binom{6+2r}{6}$ vs $\binom{2+2r}{2} + 2 \cdot \binom{4+2r}{4}$

$r = 1 \rightsquigarrow 28$ vs 36
$r = 2 \rightsquigarrow 210$ vs 155
$r = 3 \rightsquigarrow 924$ vs 448

Moment matrix

Write the first (correlative) sparse moment relaxation of

$$\begin{aligned} \inf_{\mathbf{x}} \quad & x_1 x_2 + x_1 x_3 + x_1 x_4 \\ \text{s.t.} \quad & x_1^2 + x_2^2 \leqslant 1 \\ & x_1^2 + x_3^2 \leqslant 1 \\ & x_1^2 + x_4^2 \leqslant 1 \end{aligned}$$

Moment matrix

Write the first (correlative) sparse moment relaxation of

$$\begin{aligned} \inf_{\mathbf{x}} \quad & x_1x_2 + x_1x_3 + x_1x_4 \\ \text{s.t.} \quad & x_1^2 + x_2^2 \leqslant 1 \\ & x_1^2 + x_3^2 \leqslant 1 \\ & x_1^2 + x_4^2 \leqslant 1 \end{aligned}$$

$I_1 = \{1,2\}$, $I_2 = \{1,3\}$ and $I_3 = \{1,4\}$

Moment matrix

Write the first (correlative) sparse moment relaxation of

$$\begin{aligned} \inf_{\mathbf{x}} \quad & x_1 x_2 + x_1 x_3 + x_1 x_4 \\ \text{s.t.} \quad & x_1^2 + x_2^2 \leqslant 1 \\ & x_1^2 + x_3^2 \leqslant 1 \\ & x_1^2 + x_4^2 \leqslant 1 \end{aligned}$$

$I_1 = \{1,2\}$, $I_2 = \{1,3\}$ and $I_3 = \{1,4\}$

$$\begin{aligned} \inf_{\mathbf{y}} \quad & y_{1100} + y_{1010} + y_{1001} \\ \text{s.t.} \quad & \begin{bmatrix} 1 & y_{1000} & y_{0100} \\ \star & y_{2000} & y_{1100} \\ \star & \star & y_{0200} \end{bmatrix} \succcurlyeq 0 \quad \ldots \\ & 1 - y_{2000} - y_{0200} \geqslant 0 \quad \ldots \end{aligned}$$

Measure LP preserves sparsity

$f = f_1 + f_2$, f_k depends on I_k, $\mathbf{X}$ compact & each g_j depends either on I_1 or I_2.

Measure LP preserves sparsity

$f = f_1 + f_2$, f_k depends on I_k, $\mathbf{X}$ compact & each g_j depends either on I_1 or I_2. Prove that

$$f_{\min} = \inf_{\mu \in \mathcal{M}_+(\mathbf{X})} \int_{\mathbf{X}} f \, d\mu = f_{\mathrm{cs}} = \inf_{\mu_1, \mu_2} \quad \int_{\mathbf{X}_1} f_1 \, d\mu_1 + \int_{\mathbf{X}_2} f_2 \, d\mu_2$$

$$\text{s.t.} \quad \pi_{12}\mu_1 = \pi_{21}\mu_2$$

$$\mu_1 \in \mathcal{M}_+(\mathbf{X}_1), \quad \mu_2 \in \mathcal{M}_+(\mathbf{X}_2)$$

Measure LP preserves sparsity

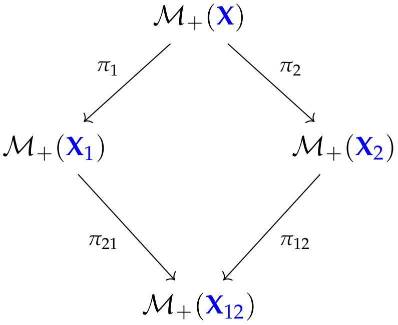
$f = f_1 + f_2$, f_k depends on I_k, $\mathbf{X}$ compact & each g_j depends either on I_1 or I_2. Prove that

$$f_{\min} = \inf_{\mu \in \mathcal{M}_+(\mathbf{X})} \int_{\mathbf{X}} f \, d\mu = f_{\text{cs}} = \inf_{\mu_1, \mu_2} \quad \int_{\mathbf{X}_1} f_1 \, d\mu_1 + \int_{\mathbf{X}_2} f_2 \, d\mu_2$$

$$\text{s.t.} \quad \pi_{12}\mu_1 = \pi_{21}\mu_2$$

$$\mu_1 \in \mathcal{M}_+(\mathbf{X}_1), \quad \mu_2 \in \mathcal{M}_+(\mathbf{X}_2)$$

(μ_k) feasible for $f_{\text{cs}} \implies \exists \mu \in \mathcal{M}_+(\mathbf{X})$ with marginal μ_k on $\mathbf{X}_k$



Measure LP preserves sparsity

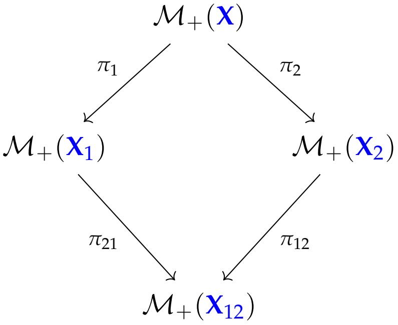
$f = f_1 + f_2$, f_k depends on I_k, $\mathbf{X}$ compact & each g_j depends either on I_1 or I_2. Prove that

$$f_{\min} = \inf_{\mu \in \mathcal{M}_+(\mathbf{X})} \int_{\mathbf{X}} f \, d\mu = f_{\mathrm{cs}} = \inf_{\mu_1, \mu_2} \quad \int_{\mathbf{X}_1} f_1 \, d\mu_1 + \int_{\mathbf{X}_2} f_2 \, d\mu_2$$

$$\text{s.t.} \quad \pi_{12}\mu_1 = \pi_{21}\mu_2$$

$$\mu_1 \in \mathcal{M}_+(\mathbf{X}_1), \quad \mu_2 \in \mathcal{M}_+(\mathbf{X}_2)$$

(μ_k) feasible for $f_{\mathrm{cs}} \implies \exists \mu \in \mathcal{M}_+(\mathbf{X})$ with marginal μ_k on $\mathbf{X}_k$



$$\boxed{f_{\mathrm{cs}} \leqslant f_{\min}}$$

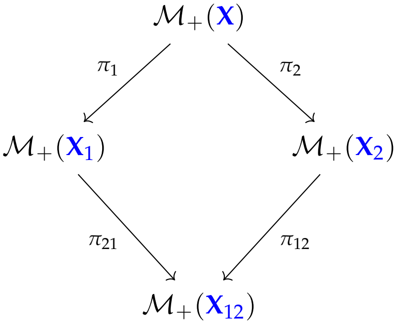
$\mu = \delta_{\mathbf{x}}$ with $\mathbf{x}$ global min. of f on $\mathbf{X}$

Measure LP preserves sparsity

$f = f_1 + f_2$, f_k depends on I_k, $\mathbf{X}$ compact & each g_j depends either on I_1 or I_2. Prove that

$$f_{\min} = \inf_{\mu \in \mathcal{M}_+(\mathbf{X})} \int_{\mathbf{X}} f \, d\mu = f_{\text{cs}} = \inf_{\mu_1, \mu_2} \quad \int_{\mathbf{X}_1} f_1 \, d\mu_1 + \int_{\mathbf{X}_2} f_2 \, d\mu_2$$
$$\text{s.t.} \quad \pi_{12}\mu_1 = \pi_{21}\mu_2$$
$$\mu_1 \in \mathcal{M}_+(\mathbf{X}_1), \quad \mu_2 \in \mathcal{M}_+(\mathbf{X}_2)$$

(μ_k) feasible for $f_{\text{cs}} \implies \exists \mu \in \mathcal{M}_+(\mathbf{X})$ with marginal μ_k on $\mathbf{X}_k$



$$\boxed{f_{\text{cs}} \leqslant f_{\min}}$$

$\mu = \delta_{\mathbf{x}}$ with $\mathbf{x}$ global min. of f on $\mathbf{X}$

$\mu_1 = \pi_1 \mu$ = Dirac at $(x_i)_{i \in I_1}$

Measure LP preserves sparsity

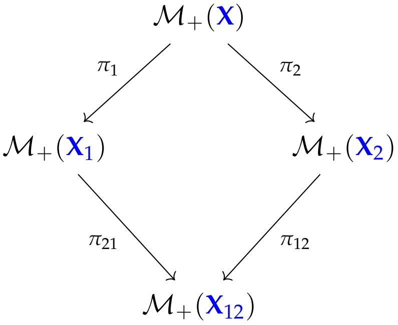
$f = f_1 + f_2$, f_k depends on I_k, $\mathbf{X}$ compact & each g_j depends either on I_1 or I_2. Prove that

$$f_{\min} = \inf_{\mu \in \mathcal{M}_+(\mathbf{X})} \int_{\mathbf{X}} f \, d\mu = f_{\text{cs}} = \inf_{\mu_1, \mu_2} \quad \int_{\mathbf{X}_1} f_1 \, d\mu_1 + \int_{\mathbf{X}_2} f_2 \, d\mu_2$$

$$\text{s.t.} \quad \pi_{12}\mu_1 = \pi_{21}\mu_2$$

$$\mu_1 \in \mathcal{M}_+(\mathbf{X}_1), \quad \mu_2 \in \mathcal{M}_+(\mathbf{X}_2)$$

(μ_k) feasible for $f_{\text{cs}} \implies \exists \mu \in \mathcal{M}_+(\mathbf{X})$ with marginal μ_k on $\mathbf{X}_k$



$f_{\text{cs}} \leqslant f_{\min}$

$\mu = \delta_{\mathbf{x}}$ with $\mathbf{x}$ global min. of f on $\mathbf{X}$

$\mu_1 = \pi_1 \mu$ = Dirac at $(x_i)_{i \in I_1}$

$f_{\text{cs}} \geqslant f_{\min}$

Measure LP preserves sparsity

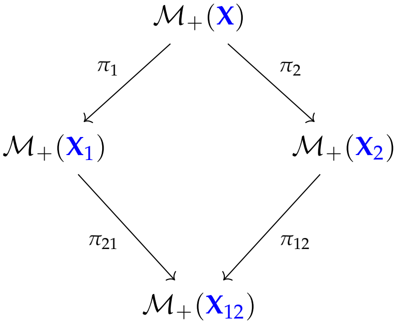
$f = f_1 + f_2$, f_k depends on I_k, $\mathbf{X}$ compact & each g_j depends either on I_1 or I_2. Prove that

$$f_{\min} = \inf_{\mu \in \mathcal{M}_+(\mathbf{X})} \int_{\mathbf{X}} f \, d\mu = f_{\text{cs}} = \inf_{\mu_1, \mu_2} \quad \int_{\mathbf{X}_1} f_1 \, d\mu_1 + \int_{\mathbf{X}_2} f_2 \, d\mu_2$$

$$\text{s.t.} \quad \pi_{12}\mu_1 = \pi_{21}\mu_2$$

$$\mu_1 \in \mathcal{M}_+(\mathbf{X}_1), \quad \mu_2 \in \mathcal{M}_+(\mathbf{X}_2)$$

(μ_k) feasible for $f_{\text{cs}} \implies \exists \mu \in \mathcal{M}_+(\mathbf{X})$ with marginal μ_k on $\mathbf{X}_k$



$$\boxed{f_{\text{cs}} \leqslant f_{\min}}$$

$\mu = \delta_{\mathbf{x}}$ with $\mathbf{x}$ global min. of f on $\mathbf{X}$

$\mu_1 = \pi_1 \mu$ = Dirac at $(x_i)_{i \in I_1}$

$$\boxed{f_{\text{cs}} \geqslant f_{\min}}$$

$$\int_{\mathbf{X}_1} f_1 \, d\mu_1 + \int_{\mathbf{X}_2} f_2 \, d\mu_2$$

$$= \int_{\mathbf{X}} f_1 \, d\mu + \int_{\mathbf{X}} f_2 \, d\mu = \int_{\mathbf{X}} f \, d\mu \geqslant f_{\min}$$

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

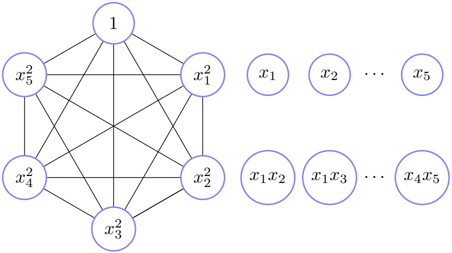
How many entries in the dense & sparse SOS/moment matrices?

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G

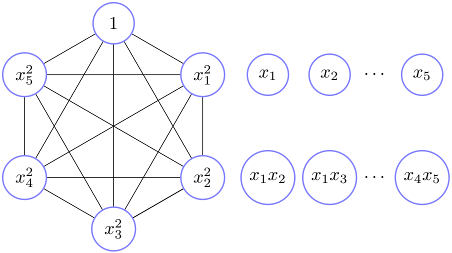


How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

💡 **tsp** graph G



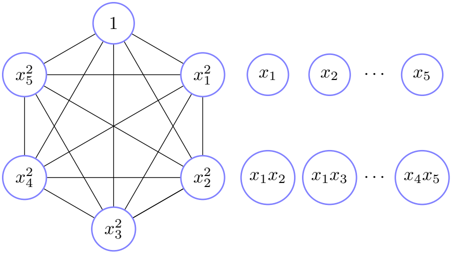
Dense = $\binom{N+2}{2}(\binom{N+2}{2}+1)/2$ in Q

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



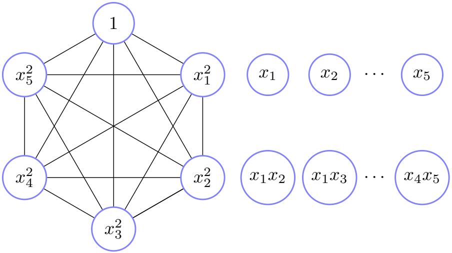
Dense = $\binom{N+2}{2}(\binom{N+2}{2}+1)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

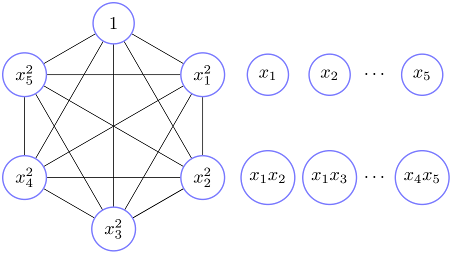
Sparse = $1 \cdot \frac{(N+1)(N+2)}{2}$

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

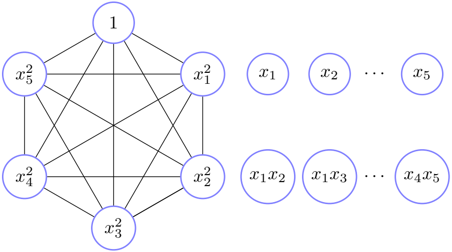
Sparse = $1 \cdot \frac{(N+1)(N+2)}{2} + N \cdot 1$

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N}(x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}(\binom{N+2}{2}+1)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

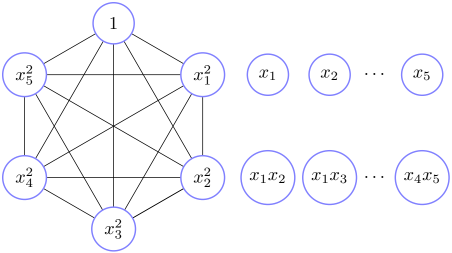
Sparse = $1 \cdot \frac{(N+1)(N+2)}{2} + N \cdot 1 + \binom{N}{2} \cdot 1$

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}(\binom{N+2}{2}+1)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

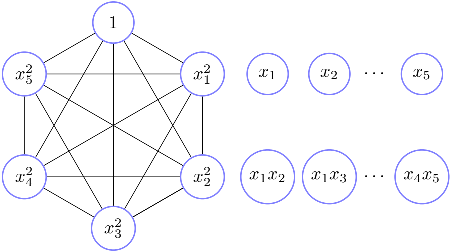
Sparse = $1 \cdot \frac{(N+1)(N+2)}{2} + N \cdot 1 + \binom{N}{2} \cdot 1 = (N+1)^2$ in Q_G

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

💡 **tsp** graph G



Dense = $\binom{N+2}{2}(\binom{N+2}{2}+1)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

Sparse = $1 \cdot \frac{(N+1)(N+2)}{2} + N \cdot 1 + \binom{N}{2} \cdot 1 = (N+1)^2$ in Q_G

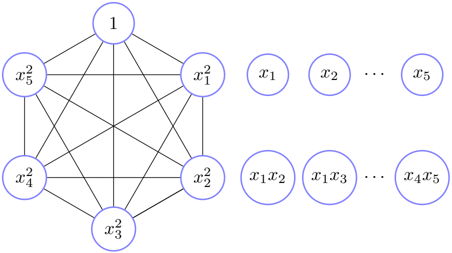
$x_i^2 \to N$

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}(\binom{N+2}{2}+1)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

Sparse = $1 \cdot \frac{(N+1)(N+2)}{2} + N \cdot 1 + \binom{N}{2} \cdot 1 = (N+1)^2$ in Q_G

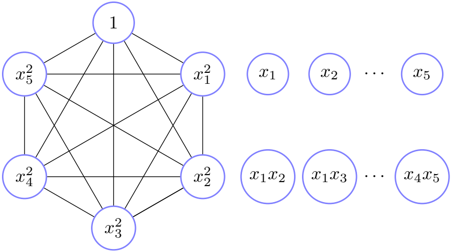
$x_i^2 \to N \quad x_i^4 \to N$

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}(\binom{N+2}{2}+1)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

Sparse = $1 \cdot \frac{(N+1)(N+2)}{2} + N \cdot 1 + \binom{N}{2} \cdot 1 = (N+1)^2$ in Q_G

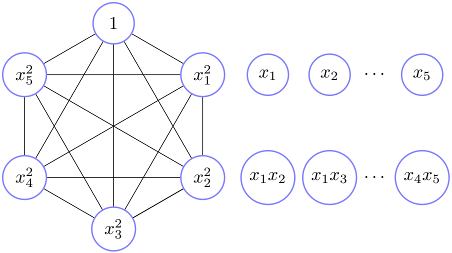
$x_i^2 \to N \quad x_i^4 \to N \quad x_i^2 x_j^2 \to \binom{N}{2}$

How big is TSSOS? (1/2)

$$f = \sum_{i=1}^{N} (x_i^2 + x_i^4) - \sum_{i,k=1}^{N} x_i^2 x_k^2$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}(\binom{N+2}{2}+1)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

Sparse = $1 \cdot \frac{(N+1)(N+2)}{2} + N \cdot 1 + \binom{N}{2} \cdot 1 = (N+1)^2$ in Q_G

$x_i^2 \to N \quad x_i^4 \to N \quad x_i^2 x_j^2 \to \binom{N}{2}$ $\binom{N}{2} + 2N$

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

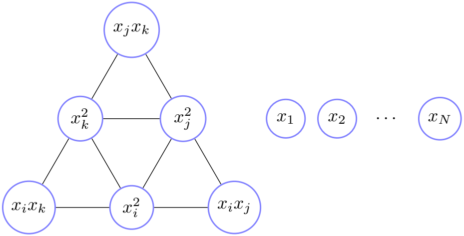
How many entries in the dense & sparse SOS/moment matrices?

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G

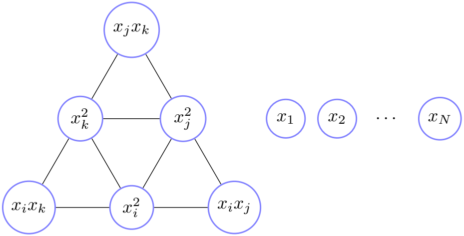


How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



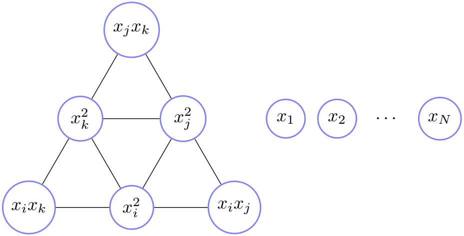
Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

💡 **tsp** graph G



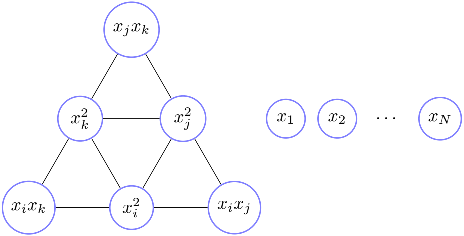
Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

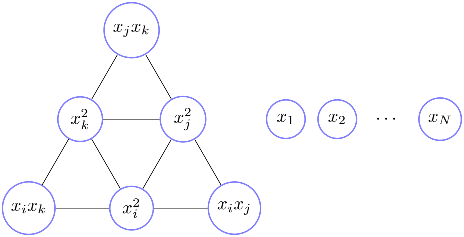
Sparse = $\frac{1 \cdot N(N+1)}{2}$

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\qquad$ $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

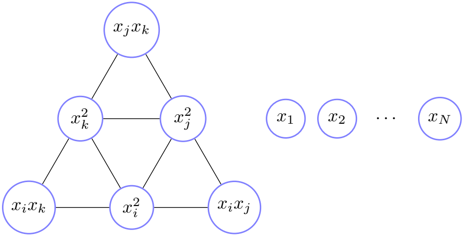
Sparse = $\frac{1 \cdot N(N+1)}{2} + \binom{N}{2} \cdot 6$

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2} + 1\right)/2$ in Q $\qquad$ $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

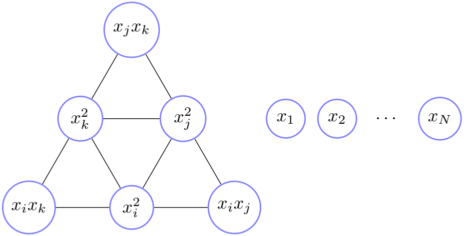
Sparse = $\frac{1 \cdot N(N+1)}{2} + \binom{N}{2} \cdot 6 + N \cdot 1$

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\qquad$ $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

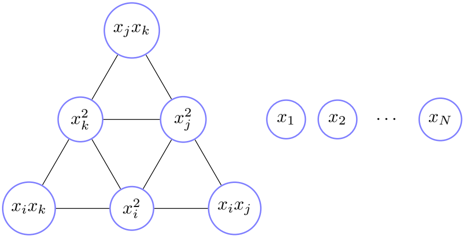
Sparse = $\frac{1 \cdot N(N+1)}{2} + \binom{N}{2} \cdot 6 + N \cdot 1$ in Q_G

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

Sparse = $\frac{1 \cdot N(N+1)}{2} + \binom{N}{2} \cdot 6 + N \cdot 1$ in Q_G

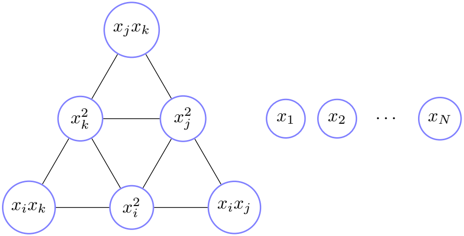
$x_i^2 \to N$

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

Sparse = $\frac{1 \cdot N(N+1)}{2} + \binom{N}{2} \cdot 6 + N \cdot 1$ in Q_G

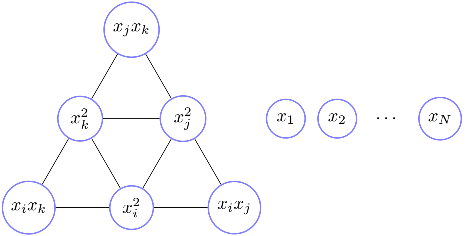
$x_i^2 \to N \quad x_i^4 \to N$

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

💡 **tsp** graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

Sparse = $\frac{1 \cdot N(N+1)}{2} + \binom{N}{2} \cdot 6 + N \cdot 1$ in Q_G

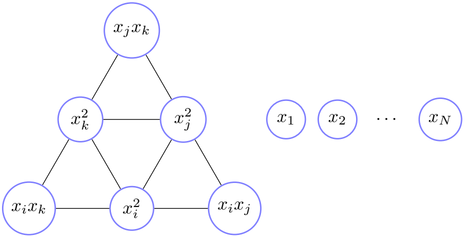
$x_i^2 \to N \quad x_i^4 \to N \quad x_i^3 x_j, x_i x_j^3, x_i^2 x_j^2 \to 3 \cdot \binom{N}{2}$

How big is TSSOS? (2/2)

$$f = \sum_{i=1}^{N} \left(\frac{x_i^2}{2} - \frac{x_i^4}{4}\right) + \sum_{i,k=1}^{N} (x_i - x_k)^4$$

How many entries in the dense & sparse SOS/moment matrices?

tsp graph G



Dense = $\binom{N+2}{2}\left(\binom{N+2}{2}+1\right)/2$ in Q $\qquad$ $\binom{N+4}{4}$ in $\mathbf{M}(\mathbf{y})$

Sparse = $\frac{1 \cdot N(N+1)}{2} + \binom{N}{2} \cdot 6 + N \cdot 1$ in Q_G

$x_i^2 \to N \quad x_i^4 \to N \quad x_i^3 x_j, x_i x_j^3, x_i^2 x_j^2 \to 3 \cdot \binom{N}{2}$ $\qquad$ $3 \cdot \binom{N}{2} + 2N$

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (1/2)

$f_1 = x_1^4 + (x_1 x_2 - 1)^2 \quad f_2 = x_2^2 x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (1/2)

$f_1 = x_1^4 + (x_1 x_2 - 1)^2 \quad f_2 = x_2^2 x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$

Compute the dense relaxation f^2
Compare with the correlative sparse relaxation f^2_{cs}
Compare with the term sparse relaxation $f^{2,s}_{\text{ts}}$ for $s = 1, 2, \ldots$

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (1/2)

$f_1 = x_1^4 + (x_1 x_2 - 1)^2 \quad f_2 = x_2^2 x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$

Compute the dense relaxation f^2
Compare with the correlative sparse relaxation f^2_{cs}
Compare with the term sparse relaxation $f^{2,s}_{\text{ts}}$ for $s = 1, 2, \ldots$

Install and run TSSOS:

```
] add https://github.com/wangjie212/TSSOS
using TSSOS, DynamicPolynomials
```

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (1/2)

$f_1 = x_1^4 + (x_1 x_2 - 1)^2 \quad f_2 = x_2^2 x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$

Compute the dense relaxation f^2
Compare with the correlative sparse relaxation f^2_{cs}
Compare with the term sparse relaxation $f^{2,s}_{\text{ts}}$ for $s = 1, 2, \ldots$

Install and run TSSOS:

```
] add https://github.com/wangjie212/TSSOS
using TSSOS, DynamicPolynomials
@polyvar x1 x2 x3; x=[x1;x2;x3];
f1 = x1^4+(x1*x2-1)^2; f2 = x2^2*x3^2+(x3^2-1)^2;
f = f1+f2
```

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (1/2)

$f_1 = x_1^4 + (x_1 x_2 - 1)^2 \quad f_2 = x_2^2 x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$

Compute the dense relaxation f^2
Compare with the correlative sparse relaxation f^2_{cs}
Compare with the term sparse relaxation $f^{2,s}_{\text{ts}}$ for $s = 1, 2, \ldots$

Install and run TSSOS:

```
] add https://github.com/wangjie212/TSSOS
using TSSOS, DynamicPolynomials
@polyvar x1 x2 x3; x=[x1;x2;x3];
f1 = x1^4+(x1*x2-1)^2; f2 = x2^2*x3^2+(x3^2-1)^2;
f = f1+f2
opt,sol,data=cs_tssos_first([f], x, 4, CS=false,TS=false);
```

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (1/2)

$f_1 = x_1^4 + (x_1 x_2 - 1)^2 \quad f_2 = x_2^2 x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$

Compute the dense relaxation f^2
Compare with the correlative sparse relaxation f^2_{cs}
Compare with the term sparse relaxation $f^{2,s}_{\text{ts}}$ for $s = 1, 2, \ldots$

Install and run TSSOS:

```
] add https://github.com/wangjie212/TSSOS
using TSSOS, DynamicPolynomials
@polyvar x1 x2 x3; x=[x1;x2;x3];
f1 = x1^4+(x1*x2-1)^2; f2 = x2^2*x3^2+(x3^2-1)^2;
f = f1+f2
opt,sol,data=cs_tssos_first([f], x, 4, CS=false,TS=false);
opt,sol,data=cs_tssos_first([f], x, 4, TS=false);
opt,sol,data=tssos_first([f], x, 4, TS="block");
opt,sol,data=tssos_higher(data, TS="block");
```

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (2/2)

Download from `https://homepages.laas.fr/vmagron/ncball`:

$f = f_1 + f_2 \qquad \mathbb{B}_{\text{nc}} = \{x : 1 - x_1^2 - x_2^2 - x_3^2 \succcurlyeq 0, 1 - x_2^2 - x_3^2 - x_4^2 \succcurlyeq 0\}$

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (2/2)

Download from `https://homepages.laas.fr/vmagron/ncball`:

$f = f_1 + f_2 \qquad \mathbb{B}_{\mathsf{nc}} = \{x : 1 - x_1^2 - x_2^2 - x_3^2 \succcurlyeq 0, 1 - x_2^2 - x_3^2 - x_4^2 \succcurlyeq 0\}$

Compute $\lambda_{\min}(f)$ on $\mathbb{B}_{\mathsf{nc}}$ with 2nd dense relaxation

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (2/2)

Download from `https://homepages.laas.fr/vmagron/ncball`:

$f = f_1 + f_2 \qquad \mathbb{B}_{\text{nc}} = \{x : 1 - x_1^2 - x_2^2 - x_3^2 \succcurlyeq 0, 1 - x_2^2 - x_3^2 - x_4^2 \succcurlyeq 0\}$

Compute $\lambda_{\min}(f)$ on $\mathbb{B}_{\text{nc}}$ with 2nd dense relaxation

```
cs_nctssos_first([f;ncball],x,2,CS=false, TS=false,
obj="eigen");
```

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (2/2)

Download from `https://homepages.laas.fr/vmagron/ncball`:

$f = f_1 + f_2 \qquad \mathbb{B}_{\mathrm{nc}} = \{x : 1 - x_1^2 - x_2^2 - x_3^2 \succcurlyeq 0, 1 - x_2^2 - x_3^2 - x_4^2 \succcurlyeq 0\}$

Compute $\lambda_{\min}(f)$ on $\mathbb{B}_{\mathrm{nc}}$ with 2nd dense relaxation

```
cs_nctssos_first([f;ncball],x,2,CS=false, TS=false,
obj="eigen");
```

Compare with the correlative and term sparse relaxations

SOS + sparse + RIP $\not\Rightarrow$ sparse SOS (2/2)

Download from `https://homepages.laas.fr/vmagron/ncball`:

$f = f_1 + f_2 \qquad \mathbb{B}_{\text{nc}} = \{x : 1 - x_1^2 - x_2^2 - x_3^2 \succcurlyeq 0, 1 - x_2^2 - x_3^2 - x_4^2 \succcurlyeq 0\}$

Compute $\lambda_{\min}(f)$ on $\mathbb{B}_{\text{nc}}$ with 2nd dense relaxation

```
cs_nctssos_first([f;ncball],x,2,CS=false, TS=false,
obj="eigen");
```

Compare with the correlative and term sparse relaxations

```
cs_nctssos_first([f;ncball],x,2,TS=false, obj="eigen");
cs_nctssos_first([f;ncball],x,3,TS=false, obj="eigen");
opt,data=nctssos_first([f;ncball],x,2,TS="MD",
obj="eigen");
opt,data = nctssos_higher!(data,TS="MD");
```