



# Exploiting sparsity in polynomial optimization

Victor Magron

## ► To cite this version:

Victor Magron. Exploiting sparsity in polynomial optimization. École thématique. Luminy, France. 2023. hal-04193936

**HAL Id: hal-04193936**

**<https://hal.science/hal-04193936>**

Submitted on 1 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

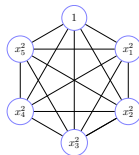
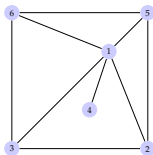
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Exploiting sparsity in polynomial optimization

Victor Magron  
LAAS CNRS

<https://homepages.laas.fr/vmagron/SparsePOPJNCF23.pdf>

Journées Nationales de Calcul Formel 6-10 March 2023



# What is sparse polynomial optimization?

---

Looks like a regular polynomial optimization problem (POP):

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

# What is sparse polynomial optimization?

---

Looks like a regular polynomial optimization problem (POP):

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

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

# What is sparse polynomial optimization?

---

Looks like a regular polynomial optimization problem (POP):

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

💡 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?

---

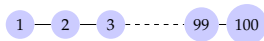
Looks like a regular polynomial optimization problem (POP):

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

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

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

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



# What is sparse polynomial optimization?

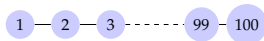
Looks like a regular polynomial optimization problem (POP):

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

💡 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?

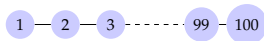
Looks like a regular polynomial optimization problem (POP):

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

💡 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  $f, g_j$

$$\rightsquigarrow f(\mathbf{x}) = x_1^{99}x_2 + x_1x_2^{100}$$

# What is sparse polynomial optimization?

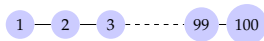
Looks like a regular polynomial optimization problem (POP):

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

💡 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  $f, g_j$

$$\rightsquigarrow f(\mathbf{x}) = x_1^{99}x_2 + x_1x_2^{100}$$

**Ideal sparsity:** constraints

# What is sparse polynomial optimization?

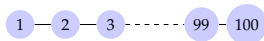
Looks like a regular polynomial optimization problem (POP):

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

💡 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  $f, g_j$

$$\rightsquigarrow f(\mathbf{x}) = x_1^{99}x_2 + x_1x_2^{100}$$

**Ideal sparsity:** constraints

$$\rightsquigarrow x_1x_2 = x_2x_3 = 0$$

# What is sparse polynomial optimization?

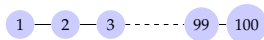
Looks like a regular polynomial optimization problem (POP):

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

💡 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  $f, g_j$

$$\rightsquigarrow f(\mathbf{x}) = x_1^{99}x_2 + x_1x_2^{100}$$

**Ideal sparsity:** constraints

$$\rightsquigarrow x_1x_2 = x_2x_3 = 0$$

PERFORMANCE



VS



ACCURACY

# Where do we find sparse POPs?

---

# Where do we find sparse POPs?

---

Everywhere (almost)!

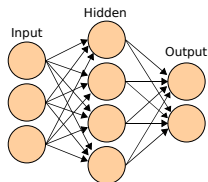
# Where do we find sparse POPs?

---

Everywhere (almost)!

## Deep learning

~> robustness, computer vision



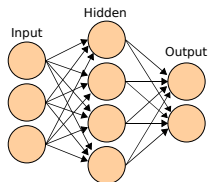
# Where do we find sparse POPs?

---

Everywhere (almost)!

## Deep learning

~> robustness, computer vision



## Power systems

~> AC optimal power-flow, stability



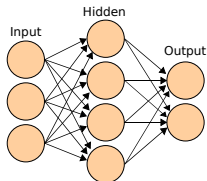
# Where do we find sparse POPs?

---

Everywhere (almost)!

## Deep learning

~> robustness, computer vision



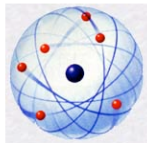
## Power systems

~> 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 $\mu$ proba $\Rightarrow$	<b>INFINITE LP</b>	$\Leftarrow$ with $f - \lambda \geq 0$

# The Moment-SOS Hierarchy for POP

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

## Practice

(Primal **Relaxation**)

moments  $\int x^\alpha d\mu$

finite number  $\Rightarrow$



**SDP**

(Dual **Strengthening**)

$f - \lambda = \text{sum of squares}$

$\Leftarrow$  fixed degree

LASSERRE'S HIERARCHY of **CONVEX PROBLEMS**  $\uparrow f_{\min}$

[Lasserre '01]

degree  $r$  &  $n$  vars  $\Rightarrow \binom{n+2r}{n}$  **SDP** VARIABLES



# The Moment-SOS Hierarchy for POP

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

## Practice

(Primal **Relaxation**)

moments  $\int x^\alpha d\mu$

finite number  $\Rightarrow$



**SDP**

(Dual **Strengthening**)

$f - \lambda = \text{sum of squares}$

$\Leftarrow$  fixed degree

LASSERRE'S HIERARCHY of **CONVEX PROBLEMS**  $\uparrow f_{\min}$

[Lasserre '01]

degree  $r$  &  $n$  vars  $\Rightarrow \binom{n+2r}{n}$  **SDP** VARIABLES



HOW TO OVERCOME THIS NO-FREE LUNCH RULE?

# An example

---

**NP hard General Problem:**  $f_{\min} := \min_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$

Semialgebraic set  $\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0\}$

# An example

---

**NP hard General Problem:**  $f_{\min} := \min_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$

Semialgebraic set  $\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0\}$

$$\mathbf{X} = [0, 1]^2 = \{\mathbf{x} \in \mathbb{R}^2 : x_1(1 - x_1) \geq 0, \quad x_2(1 - x_2) \geq 0\}$$

# An example

---

**NP hard General Problem:**  $f_{\min} := \min_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$

Semialgebraic set  $\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0\}$

$$\mathbf{X} = [0, 1]^2 = \{\mathbf{x} \in \mathbb{R}^2 : x_1(1 - x_1) \geq 0, \quad x_2(1 - x_2) \geq 0\}$$

$$\overbrace{x_1 x_2}^f = -\frac{1}{8} + \overbrace{\frac{1}{2} \left( x_1 + x_2 - \frac{1}{2} \right)^2}^{\sigma_0} + \overbrace{\frac{1}{2} x_1(1 - x_1)}^{\sigma_1 \overbrace{x_1(1 - x_1)}^{g_1}} + \overbrace{\frac{1}{2} x_2(1 - x_2)}^{\sigma_2 \overbrace{x_2(1 - x_2)}^{g_2}}$$

# An example

---

**NP hard General Problem:**  $f_{\min} := \min_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$

Semialgebraic set  $\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0\}$

$$\mathbf{X} = [0, 1]^2 = \{\mathbf{x} \in \mathbb{R}^2 : x_1(1 - x_1) \geq 0, \quad x_2(1 - x_2) \geq 0\}$$

$$\overbrace{x_1 x_2}^f = -\frac{1}{8} + \overbrace{\frac{1}{2} \left( x_1 + x_2 - \frac{1}{2} \right)^2}^{\sigma_0} + \overbrace{\frac{1}{2} x_1(1 - x_1)}^{\sigma_1 \overbrace{x_1(1 - x_1)}^{g_1}} + \overbrace{\frac{1}{2} x_2(1 - x_2)}^{\sigma_2 \overbrace{x_2(1 - x_2)}^{g_2}}$$

Sums of squares (SOS)  $\sigma_j$

# An example

---

**NP hard General Problem:**  $f_{\min} := \min_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x})$

Semialgebraic set  $\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0\}$

$$\mathbf{X} = [0, 1]^2 = \{\mathbf{x} \in \mathbb{R}^2 : x_1(1 - x_1) \geq 0, \quad x_2(1 - x_2) \geq 0\}$$

$$\overbrace{x_1 x_2}^f = -\frac{1}{8} + \frac{1}{2} \overbrace{\left(x_1 + x_2 - \frac{1}{2}\right)^2}^{\sigma_0} + \frac{1}{2} \overbrace{x_1(1 - x_1)}^{\sigma_1 \overbrace{g_1}^{g_1}} + \frac{1}{2} \overbrace{x_2(1 - x_2)}^{\sigma_2 \overbrace{g_2}^{g_2}}$$

Sums of squares (SOS)  $\sigma_j$

$$\text{Quadratic module: } \mathcal{M}(\mathbf{X})_r = \left\{ \sigma_0 + \sum_j \sigma_j g_j, \deg \sigma_j g_j \leq 2r \right\}$$

# 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)
$\inf \int_{\mathbf{X}} f d\mu$	$=$	$\sup \lambda$
s.t. $\mu \in \mathcal{M}_+(\mathbf{X})$		s.t. $\lambda \in \mathbb{R}$
		$f - \lambda \in \mathcal{Q}(\mathbf{X})$

# 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. $\lambda \in \mathbb{R}$
$y_0 = 1$		$f - \lambda \in \mathcal{Q}(\mathbf{X})_r$

# 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

Term sparsity

Ideal sparsity

Tutorial session

Sparse SDP

Correlative sparsity

Term sparsity

Ideal sparsity

Tutorial session

# Sparse matrices

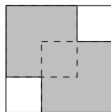
---

Symmetric matrices indexed by graph vertices

# Sparse matrices

---

Symmetric matrices indexed by graph vertices

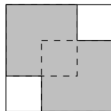


💡 no edge between 1 and 3  $\iff$  0 entry in the (1,3) entry

# Sparse matrices

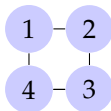
---

Symmetric matrices indexed by graph vertices



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

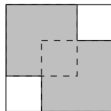
cycle =



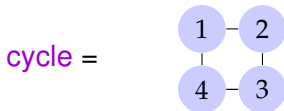
# Sparse matrices

---

Symmetric matrices indexed by graph vertices



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

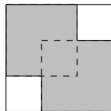


chord = edge between two nonconsecutive vertices in a cycle

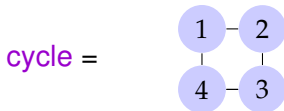
# Sparse matrices

---

Symmetric matrices indexed by graph vertices

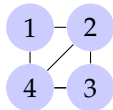


💡 no edge between 1 and 3  $\iff$  0 entry in the (1,3) entry



**chord** = edge between two nonconsecutive vertices in a cycle

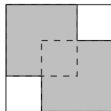
**chordal** graph = all cycles of length  $\geq 4$  have at least one chord



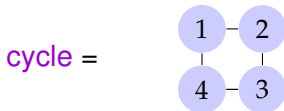
# Sparse matrices

---

Symmetric matrices indexed by graph vertices



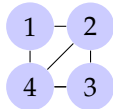
💡 no edge between 1 and 3  $\iff$  0 entry in the (1,3) entry



**chord** = edge between two nonconsecutive vertices in a cycle

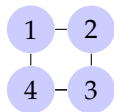
**chordal** graph = all cycles of length  $\geq 4$  have at least one chord

**clique** = a fully connected subset of vertices



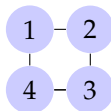
# Chordal extensions

---



# Chordal extensions

---

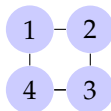


## Fact

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

# Chordal extensions

---

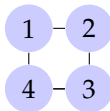


## Fact

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

💡 Chordal extension is not unique!

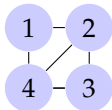
# Chordal extensions



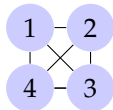
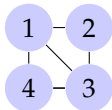
## Fact

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  $I_1, \dots, I_p$ :

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

(possibly after reordering)

# Running intersection property (RIP)

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

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

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

(possibly after reordering)

💡 RIP always holds for  $p = 2$

# Running intersection property (RIP)

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

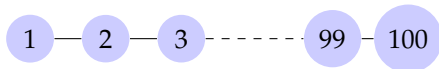
For a chordal graph with maximal cliques  $I_1, \dots, I_p$ :

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

(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]

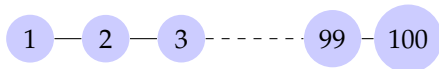
For a chordal graph with maximal cliques  $I_1, \dots, I_p$ :

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

(possibly after reordering)

💡 RIP always holds for  $p = 2$

💡 RIP holds for chains



💡 RIP holds for numerous applications!

# Semidefinite Programming (SDP)

---

$$\begin{array}{ll}\min_{\mathbf{y}} & \mathbf{c}^\top \mathbf{y} \\ \text{s.t.} & \sum_i \mathbf{F}_i y_i \succcurlyeq \mathbf{F}_0\end{array}$$

- 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**

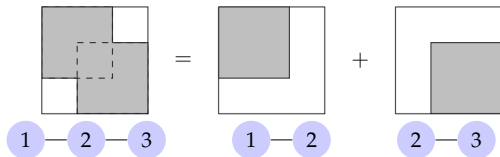
# Sparse SDP matrices

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  $G$

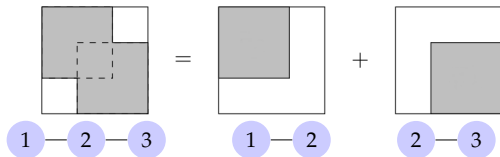
$\implies Q_G = P_1^T Q_1 P_1 + P_2^T Q_2 P_2$  with  $Q_k \succcurlyeq 0$  indexed by  $I_k$



# 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  $P_1, P_2$ ?

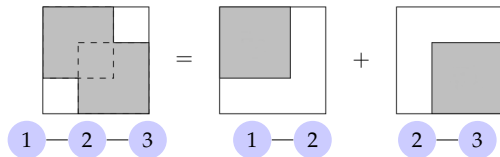
# Sparse SDP matrices

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  $G$

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



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

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

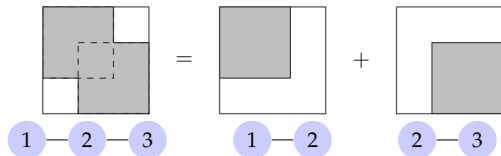
# Sparse SDP matrices

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  $G$

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



What are  $P_1, P_2$ ?  $P_1 \in \mathbb{R}^{|I_1| \times n}$   $P(i, j) = \begin{cases} 1 & \text{if } I(i) = j \\ 0 & \text{otherwise} \end{cases}$

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

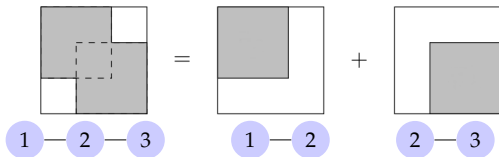
# Sparse SDP matrices

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  $G$

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



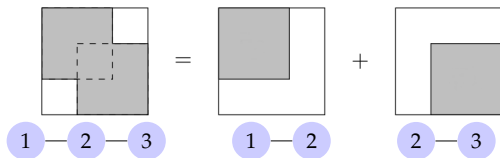
What are  $P_1, P_2$ ?  $P_1 \in \mathbb{R}^{|I_1| \times n}$   $P(i, j) = \begin{cases} 1 & \text{if } I(i) = j \\ 0 & \text{otherwise} \end{cases}$

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

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



What are  $P_1, P_2$ ?  $P_1 \in \mathbb{R}^{|I_1| \times n}$   $P(i, j) = \begin{cases} 1 & \text{if } I(i) = j \\ 0 & \text{otherwise} \end{cases}$

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

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

Sparse SDP

**Correlative sparsity**

Term sparsity

Ideal sparsity

Tutorial session

# What is correlative sparsity?

---

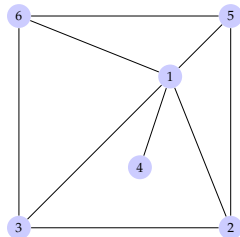
💡 Exploit few links between **variables** [Lasserre, Waki et al. '06]

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

Correlative sparsity pattern (csp) graph  $G$

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

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



# What is correlative sparsity?

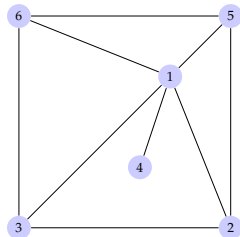
💡 Exploit few links between **variables** [Lasserre, Waki et al. '06]

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

Correlative sparsity pattern (csp) graph  $G$

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

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



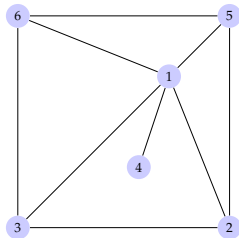
Similar construction with constraints  $\mathbf{X} = \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0\}$

# 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)

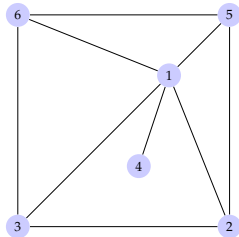


# What is correlative sparsity?

---

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

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  $N - \sum_i x_i^2 \geq 0$  then

$f > 0$  on  $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0\} \implies f = \sigma_0 + \sum_j \sigma_j g_j$  with  $\sigma_j$  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  $N - \sum_i x_i^2 \geq 0$  then

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

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

$\implies$  SOS  $\sigma_{0k}$  "sees" vars in  $I_k$   
 $\sigma_j$  "sees" vars from  $g_j$

# A first key message

---



**SUMS OF SQUARES PRESERVE SPARSITY**



# A proof of sparse Putinar's Positivstellensatz

---

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

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

# A proof of sparse Putinar's Positivstellensatz

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

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geq 0 : j \in J_k\}$  = 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  $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0\}$  be compact and  $f = \sum_k f_k$ , with  $f_k$  depends on  $\mathbf{x}(I_k)$ , and  $f > 0$  on  $\mathbf{X}$

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geq 0 : j \in J_k\}$  = 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  $\mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0\}$  be compact and  $f = \sum_k f_k$ , with  $f_k$  depends on  $\mathbf{x}(I_k)$ , and  $f > 0$  on  $\mathbf{X}$

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geq 0 : j \in J_k\}$  = 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  $\mathbf{x}(I_k)$

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

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

$$\mathbf{M}_1(\mathbf{y}, I_1) = \left( \begin{array}{c|cc} 1 & y_{1,0,0,0,0,0} & y_{0,0,0,1,0,0} \\ \hline y_{1,0,0,0,0,0} & y_{2,0,0,0,0,0} & y_{1,0,0,1,0,0} \\ y_{0,0,0,1,0,0} & y_{1,0,0,1,0,0} & y_{0,0,0,2,0,0} \end{array} \right)$$

# Sparse moment matrices

---

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

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

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

$$\mathbf{M}_1(\mathbf{y}, I_1) = \left( \begin{array}{c|cc} 1 & y_{1,0,0,0,0,0} & y_{0,0,0,1,0,0} \\ \hline y_{1,0,0,0,0,0} & y_{2,0,0,0,0,0} & y_{1,0,0,1,0,0} \\ y_{0,0,0,1,0,0} & y_{1,0,0,1,0,0} & y_{0,0,0,2,0,0} \end{array} \right)$$

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

# Sparse primal-dual Moment-SOS hierarchy

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

## Dense Moment-SOS hierarchy

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

# Sparse primal-dual Moment-SOS hierarchy

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

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

Each  $g_j$  depends on some  $I_k$

## Sparse Moment-SOS hierarchy

(Moment)

$$\inf \sum_{\alpha} f_{\alpha} y_{\alpha}$$

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

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

$$y_0 = 1$$

(SOS)

$$= \sup \lambda$$

$$\text{s.t. } \lambda \in \mathbb{R}$$

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

# Sparse primal-dual Moment-SOS hierarchy

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

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

Each  $g_j$  depends on some  $I_k$

## Sparse Moment-SOS hierarchy

(Moment)

$$\inf \sum_{\alpha} f_{\alpha} y_{\alpha}$$

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

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

$$y_0 = 1$$

(SOS)

$$= \sup \lambda$$

$$\text{s.t. } \lambda \in \mathbb{R}$$

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

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

$$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x}) \text{ with } \mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, j \leq m\}$$

$$\tau = \max\{|I_1|, \dots, |I_p|\}$$

## Sparse Moment-SOS hierarchy

(Moment)

$$\inf \sum_{\alpha} f_{\alpha} y_{\alpha}$$

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

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

$$y_0 = 1$$

(SOS)

$$= \sup \lambda$$

$$\text{s.t. } \lambda \in \mathbb{R}$$

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

# Computational cost

$$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x}) \text{ with } \mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, j \leq m\}$$

$$\tau = \max\{|I_1|, \dots, |I_p|\}$$

## Sparse Moment-SOS hierarchy

(Moment)

$$\inf \sum_{\alpha} f_{\alpha} y_{\alpha}$$

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

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

$$y_0 = 1$$

(SOS)

$$= \sup \lambda$$

$$\text{s.t. } \lambda \in \mathbb{R}$$

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

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

# Computational cost

$$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x}) \text{ with } \mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, j \leq m\}$$

$$\tau = \max\{|I_1|, \dots, |I_p|\}$$

## Sparse Moment-SOS hierarchy

(Moment)

$$\inf \sum_{\alpha} f_{\alpha} y_{\alpha}$$

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

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

$$y_0 = 1$$

(SOS)

$$= \sup \lambda$$

$$\text{s.t. } \lambda \in \mathbb{R}$$

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

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

💡  $(m + p) \mathcal{O}(r^{\tau})$  SDP vars

# Computational cost

$$f_{\min} = \inf_{\mathbf{x} \in \mathbf{X}} f(\mathbf{x}) \text{ with } \mathbf{X} = \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, j \leq m\}$$

$$\tau = \max\{|I_1|, \dots, |I_p|\}$$

## Sparse Moment-SOS hierarchy

(Moment)

$$\inf \sum_{\alpha} f_{\alpha} y_{\alpha}$$

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

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

$$y_0 = 1$$

(SOS)

$$= \sup \lambda$$

$$\text{s.t. } \lambda \in \mathbb{R}$$

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

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

💡  $(m + p) \mathcal{O}(r^{\tau})$  SDP vars vs  $(m + 1) \mathcal{O}(r^n)$  in the dense SDP

# Sparse linear program over measures

---

In the dense setting:

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

# Sparse linear program over measures

---

In the dense setting:

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

In the sparse setting:

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

# Sparse linear program over measures

---

In the dense setting:

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

In the sparse setting:

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geq 0 : j \in J_k\}$  = 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:

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

In the sparse setting:

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geq 0 : j \in J_k\}$  = 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  $\mu_k$  for each  $I_k \rightarrow$  marginals  $\pi_{kj}\mu_k$  on  $\mathcal{M}_+(\mathbf{X}_{kj})$

# Sparse linear program over measures

---

In the dense setting:

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

In the sparse setting:

$\mathbf{X}_k = \{\mathbf{x}(I_k) : g_j(\mathbf{x}) \geq 0 : j \in J_k\}$  = 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  $\mu_k$  for each  $I_k \rightarrow$  marginals  $\pi_{kj}\mu_k$  on  $\mathcal{M}_+(\mathbf{X}_{kj})$

Sparse moment SDPs relax the sparse LP over measures:

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

# The dual of sparse Putinar's Positivstellensatz

## Theorem [Lasserre '06]

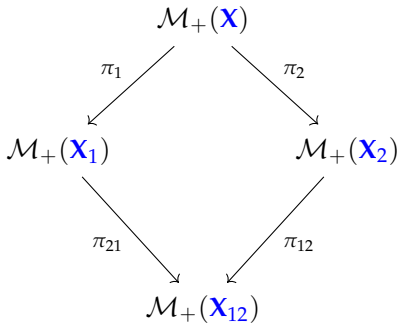
$$\begin{aligned} \text{RIP holds for } (I_k) \implies f_{\min} = f_{\text{cs}} = \inf_{\mu_k} \quad & \sum_k \int_{\mathbf{X}_k} f_k d\mu_k \\ \text{s.t.} \quad & \pi_{jk} \mu_j = \pi_{kj} \mu_k \\ & \mu_k \in \mathcal{M}_+(\mathbf{X}_k) \end{aligned}$$

# The dual of sparse Putinar's Positivstellensatz

## Theorem [Lasserre '06]

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

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

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

# 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

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

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

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

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

Extraction possible with the Gloptipoly software

# Extracting minimizers: the sparse case

---

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

# Extracting minimizers: the sparse case

---

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

**Theorem: sparse extraction [Lasserre '06]**

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

$$\text{rank } \mathbf{M}_r(\mathbf{y}, I_k) = \text{rank } \mathbf{M}_{r-a_k}(\mathbf{y}, I_k)$$

$$\text{rank } \mathbf{M}_r(\mathbf{y}, I_k \cap I_j) = 1$$

# 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  $f_{\text{CS}}^r$  and

$$\text{rank } \mathbf{M}_r(\mathbf{y}, I_k) = \text{rank } \mathbf{M}_{r-a_k}(\mathbf{y}, I_k)$$

$$\text{rank } \mathbf{M}_r(\mathbf{y}, I_k \cap I_j) = 1$$

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

# Extracting minimizers: the sparse case

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

**Theorem: sparse extraction [Lasserre '06]**

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

$$\text{rank } \mathbf{M}_r(\mathbf{y}, I_k) = \text{rank } \mathbf{M}_{r-a_k}(\mathbf{y}, I_k)$$

$$\text{rank } \mathbf{M}_r(\mathbf{y}, I_k \cap I_j) = 1$$

Then  $f_{\text{CS}}^r = f_{\min} = f_{\text{CS}}$  & sparse measure LP has optimal solution  $\mu_k \in \mathcal{M}_+(\mathbf{X}_k)$  supported on  $t_k = \text{rank } \mathbf{M}_r(\mathbf{y}, I_k)$  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  $f(\mathbf{x}) = x_1x_2 + x_3x_4$

# Application to roundoff errors

---

[Magron Constantinides Donaldson '17]

Exact  $f(\mathbf{x}) = x_1x_2 + x_3x_4$

Floating-point  $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1x_2(1 + e_1) + x_3x_4(1 + e_2)](1 + e_3)$

# Application to roundoff errors

---

[Magron Constantinides Donaldson '17]

Exact  $f(\mathbf{x}) = x_1x_2 + x_3x_4$

Floating-point  $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1x_2(1 + e_1) + x_3x_4(1 + e_2)](1 + e_3)$

$\mathbf{x} \in \mathbf{X}$ ,  $|e_i| \leq 2^{-\delta}$   $\delta = 24$  (single) or 53 (double)

# Application to roundoff errors

---

[Magron Constantinides Donaldson '17]

Exact  $f(\mathbf{x}) = x_1x_2 + x_3x_4$

Floating-point  $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1x_2(1 + e_1) + x_3x_4(1 + e_2)](1 + e_3)$

$\mathbf{x} \in \mathbf{X}$ ,  $|e_i| \leq 2^{-\delta}$   $\delta = 24$  (single) or 53 (double)

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

# Application to roundoff errors

---

[Magron Constantinides Donaldson '17]

Exact  $f(\mathbf{x}) = x_1x_2 + x_3x_4$

Floating-point  $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1x_2(1 + e_1) + x_3x_4(1 + e_2)](1 + e_3)$

$\mathbf{x} \in \mathbf{X}$ ,  $|e_i| \leq 2^{-\delta}$   $\delta = 24$  (single) or 53 (double)

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

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

# Application to roundoff errors

---

[Magron Constantinides Donaldson '17]

Exact  $f(\mathbf{x}) = x_1x_2 + x_3x_4$

Floating-point  $\hat{f}(\mathbf{x}, \mathbf{e}) = [x_1x_2(1 + e_1) + x_3x_4(1 + e_2)](1 + e_3)$

$\mathbf{x} \in \mathbf{X}$ ,  $|e_i| \leq 2^{-\delta}$   $\delta = 24$  (single) or 53 (double)

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

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

# 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

---

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

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

**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:**  $922\epsilon$  ( $10 \times$  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  $\leadsto$  **Out of memory**

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

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

**Symbolic Taylor** FPTaylor tool:  $721\epsilon$  ( $21 \times$  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:  $15\binom{6+1+4}{6+1} = 4950 \leadsto 759\epsilon$

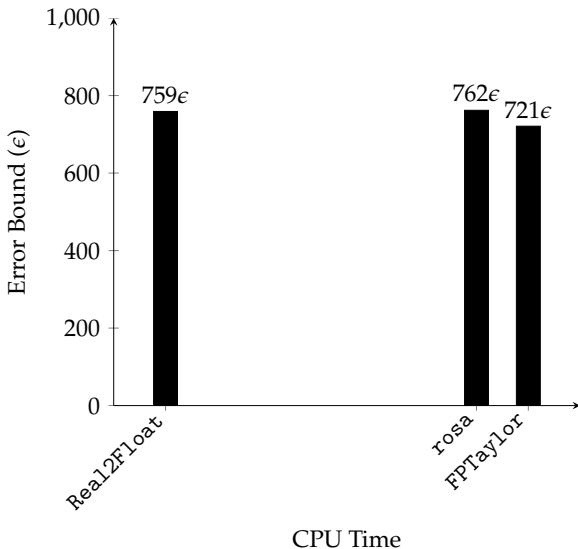
**Interval arithmetic:**  $922\epsilon$  ( $10 \times$  less CPU)

**Symbolic Taylor** FPTaylor tool:  $721\epsilon$  ( $21 \times$  more CPU)

**SMT-based** rosa tool:  $762\epsilon$  ( $19 \times$  more CPU)

# Application to roundoff errors

---



# Extension to noncommutative optimization

---

Self-adjoint noncommutative variables  $a_i, b_j \in \mathcal{B}(\mathcal{H})$

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - b_1 - 2b_1 - b_2$$

with  $a_1a_2 \neq a_2a_1$ , **involution**  $(a_1b_3)^* = b_3a_1$

# Extension to noncommutative optimization

---

Self-adjoint noncommutative variables  $a_i, b_j \in \mathcal{B}(\mathcal{H})$

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - b_1 - 2b_1 - b_2$$

with  $a_1a_2 \neq a_2a_1$ , **involution**  $(a_1b_3)^* = b_3a_1$

Constraints  $\mathbf{X} = \{(a, b) : a_i^2 = a_i, b_j^2 = b_j, a_ib_j = b_ja_i\}$

# Extension to noncommutative optimization

---

Self-adjoint noncommutative variables  $a_i, b_j \in \mathcal{B}(\mathcal{H})$

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - b_1 - 2b_1 - b_2$$

with  $a_1a_2 \neq a_2a_1$ , **involution**  $(a_1b_3)^* = b_3a_1$

Constraints  $\mathbf{X} = \{(a, b) : a_i^2 = a_i, b_j^2 = b_j, a_ib_j = b_ja_i\}$

## MINIMAL EIGENVALUE OPTIMIZATION

$$\lambda_{\min} = \inf \{ \langle f(a, b) \mathbf{v}, \mathbf{v} \rangle : (a, b) \in \mathbf{X}, \|\mathbf{v}\| = 1 \}$$

# Extension to noncommutative optimization

---

Self-adjoint noncommutative variables  $a_i, b_j \in \mathcal{B}(\mathcal{H})$

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - b_1 - 2b_1 - b_2$$

with  $a_1 a_2 \neq a_2 a_1$ , **involution**  $(a_1 b_3)^* = b_3 a_1$

Constraints  $\mathbf{X} = \{(a, b) : a_i^2 = a_i, b_j^2 = b_j, a_i b_j = b_j a_i\}$

## MINIMAL EIGENVALUE OPTIMIZATION

$$\begin{aligned}\lambda_{\min} &= \inf \{ \langle f(a, b) \mathbf{v}, \mathbf{v} \rangle : (a, b) \in \mathbf{X}, \|\mathbf{v}\| = 1 \} \\ &= \sup \quad \lambda \\ \text{s.t.} \quad & f(a, b) - \lambda \mathbf{I} \succcurlyeq 0, \quad \forall (a, b) \in \mathbf{X}\end{aligned}$$

# Extension to noncommutative optimization

Ball constraint  $N - \sum_i x_i^2 \succcurlyeq 0$  in  $\mathbf{X}$

Theorem: NC Putinar's representation [Helton & McCullough '02]

$$f \succcurlyeq 0 \text{ on } \mathbf{X} \implies f = \sum_i s_i^* s_i + \sum_j \sum_i t_{ji}^* g_j t_{ji} \text{ with } s_i, t_{ji} \in \mathbb{R}\langle \underline{x} \rangle$$

# Extension to noncommutative optimization

Ball constraint  $N - \sum_i x_i^2 \succcurlyeq 0$  in  $\mathbf{X}$

Theorem: NC Putinar's representation [Helton & McCullough '02]

$$f \succcurlyeq 0 \text{ on } \mathbf{X} \implies \boxed{f = \sum_i s_i^* s_i + \sum_j \sum_i t_{ji}^* g_j t_{ji}} \text{ with } s_i, t_{ji} \in \mathbb{R}\langle \underline{x} \rangle$$

NC variant of Lasserre's Hierarchy for  $\lambda_{\min}$ :

💡 replace “ $f - \lambda \mathbf{I} \succcurlyeq 0$  on  $\mathbf{X}$ ” by  $f - \lambda \mathbf{I} = \sum_i s_i^* s_i + \sum_j \sum_i t_{ji}^* g_j t_{ji}$   
with  $s_i, t_{ji}$  of **bounded** degrees = SDP 🎯

# Extension to noncommutative optimization

---

Self-adjoint noncommutative (NC) variables  $x = (x_1, \dots, x_n)$

Theorem [Helton & McCullough '02]

$f \succcurlyeq 0 \Leftrightarrow f$  SOS (all positive polynomials are sums of squares)

# Extension to noncommutative optimization

---

Self-adjoint noncommutative (NC) variables  $x = (x_1, \dots, x_n)$

Theorem [Helton & McCullough '02]

$f \succcurlyeq 0 \Leftrightarrow f$  SOS (all positive polynomials are sums of squares)

**BAD NEWS:** there is **no** sparse analog!

[Klep Magron Povh '21]

sparse  $f$  SOS  $\nRightarrow f$  is a sparse SOS

# Extension to noncommutative optimization

Self-adjoint noncommutative (NC) variables  $x = (x_1, \dots, x_n)$

Theorem [Helton & McCullough '02]

$f \succcurlyeq 0 \Leftrightarrow f$  SOS (all positive polynomials are sums of squares)

**BAD NEWS:** there is **no** sparse analog!

sparse  $f$  SOS  $\nRightarrow f$  is a sparse SOS

[Klep Magron Povh '21]

Take  $f = (x_1 + x_2 + x_3)^2$

# Extension to noncommutative optimization

Self-adjoint noncommutative (NC) variables  $x = (x_1, \dots, x_n)$

Theorem [Helton & McCullough '02]

$f \succcurlyeq 0 \Leftrightarrow f$  SOS (all positive polynomials are sums of squares)

**BAD NEWS:** there is **no** sparse analog!

sparse  $f$  SOS  $\not\Rightarrow f$  is a sparse SOS

[Klep Magron Povh '21]

Take  $f = (x_1 + x_2 + x_3)^2$

**GOOD NEWS:** there is an NC analog of the sparse Putinar's Positivstellensatz!

# Extension to noncommutative optimization

Self-adjoint noncommutative (NC) variables  $x = (x_1, \dots, x_n)$

Theorem [Helton & McCullough '02]

$f \succcurlyeq 0 \Leftrightarrow f$  SOS (all positive polynomials are sums of squares)

**BAD NEWS:** there is **no** sparse analog!

sparse  $f$  SOS  $\not\Rightarrow f$  is a sparse SOS

[Klep Magron Povh '21]

Take  $f = (x_1 + x_2 + x_3)^2$

**GOOD NEWS:** there is an NC analog of the sparse Putinar's Positivstellensatz! Based on GNS construction & **amalgamation**  
[Blackadar '78, Voiculescu '85]

# Extension to noncommutative optimization

Self-adjoint noncommutative (NC) variables  $x = (x_1, \dots, x_n)$

Theorem [Helton & McCullough '02]

$f \succcurlyeq 0 \Leftrightarrow f$  SOS (all positive polynomials are sums of squares)

**BAD NEWS:** there is **no** sparse analog!

sparse  $f$  SOS  $\nRightarrow f$  is a sparse SOS

[Klep Magron Povh '21]

Take  $f = (x_1 + x_2 + x_3)^2$

**GOOD NEWS:** there is an NC analog of the sparse Putinar's Positivstellensatz! Based on GNS construction & **amalgamation**

[Blackadar '78, Voiculescu '85]

Theorem: Sparse NC Positivstellensatz [Klep Magron Povh '21]

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

$f > 0$  on  $\{x : g_j(x) \geq 0\}$

chordal graph with cliques  $I_k \Rightarrow$

ball constraints for each  $x(I_k)$

$$f = \sum_{k,i} (s_{ki}^* s_{ki} + \sum_{j \in I_k} t_{ji}^* g_j t_{ji})$$

$s_{ki}$  "sees" vars in  $I_k$

$t_{ji}$  "sees" vars from  $g_j$

# Application to violation of Bell inequalities

---

I<sub>3322</sub> Bell inequality (entanglement in quantum information)

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - a_1 - 2b_1 - b_2$$

Maximal violation levels  $\rightarrow$  **upper bounds** on  $\lambda_{\max}$  of  $f$  on  $\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$

# Application to violation of Bell inequalities

---

I<sub>3322</sub> Bell inequality (entanglement in quantum information)

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - a_1 - 2b_1 - b_2$$

Maximal violation levels  $\rightarrow$  **upper bounds** on  $\lambda_{\max}$  of  $f$  on

$$\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$$

💡  $I_k \rightarrow \{a_k, b_1, b_2, b_3\}$

# Application to violation of Bell inequalities

---

I<sub>3322</sub> Bell inequality (entanglement in quantum information)

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - a_1 - 2b_1 - b_2$$

Maximal violation levels  $\rightarrow$  **upper bounds** on  $\lambda_{\max}$  of  $f$  on

$$\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$$

💡  $I_k \rightarrow \{a_k, b_1, b_2, b_3\}$

level	sparse	dense [Pál & Vértesi '18]
2	0.2550008	0.2509397

# Application to violation of Bell inequalities

---

I<sub>3322</sub> Bell inequality (entanglement in quantum information)

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - a_1 - 2b_1 - b_2$$

Maximal violation levels  $\rightarrow$  **upper bounds** on  $\lambda_{\max}$  of  $f$  on

$$\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$$

💡  $I_k \rightarrow \{a_k, b_1, b_2, b_3\}$

level	sparse	dense [Pál & Vértesi '18]
2	0.2550008	0.2509397
3	0.2511592	0.2508756

# Application to violation of Bell inequalities

---

I<sub>3322</sub> Bell inequality (entanglement in quantum information)

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - a_1 - 2b_1 - b_2$$

Maximal violation levels  $\rightarrow$  **upper bounds** on  $\lambda_{\max}$  of  $f$  on

$$\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$$

💡  $I_k \rightarrow \{a_k, b_1, b_2, b_3\}$

level	sparse	dense [Pál & Vértesi '18]
2	0.2550008	0.2509397
3	0.2511592	0.2508756
3'		0.2508754 (1 day)

# Application to violation of Bell inequalities

---

I<sub>3322</sub> Bell inequality (entanglement in quantum information)

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - a_1 - 2b_1 - b_2$$

Maximal violation levels  $\rightarrow$  **upper bounds** on  $\lambda_{\max}$  of  $f$  on

$$\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$$

💡  $I_k \rightarrow \{a_k, b_1, b_2, b_3\}$

level	sparse	dense [Pál & Vértesi '18]
2	0.2550008	0.2509397
3	0.2511592	0.2508756
3'		0.2508754 (1 day)
4	0.2508917	

# Application to violation of Bell inequalities

I<sub>3322</sub> Bell inequality (entanglement in quantum information)

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - a_1 - 2b_1 - b_2$$

Maximal violation levels  $\rightarrow$  **upper bounds** on  $\lambda_{\max}$  of  $f$  on

$$\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$$

💡  $I_k \rightarrow \{a_k, b_1, b_2, b_3\}$

level	sparse	dense [Pál & Vértesi '18]
2	0.2550008	0.2509397
3	0.2511592	0.2508756
3'		0.2508754 (1 day)
4	0.2508917	
5	0.2508763	

# Application to violation of Bell inequalities

I<sub>3322</sub> Bell inequality (entanglement in quantum information)

$$f = a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 - b_3) + a_3(b_1 - b_2) - a_1 - 2b_1 - b_2$$

Maximal violation levels  $\rightarrow$  **upper bounds** on  $\lambda_{\max}$  of  $f$  on  $\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$

💡  $I_k \rightarrow \{a_k, b_1, b_2, b_3\}$

level	sparse	dense [Pál & Vértesi '18]
2	0.2550008	0.2509397
3	0.2511592	0.2508756
3'		0.2508754 (1 day)
4	0.2508917	
5	0.2508763	
6	0.2508753977180 (1 hour)	

PERFORMANCE



VS



ACCURACY

# Application to SOS of bounded degrees

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

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

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

with  $\sigma_k$  SOS of degree  $\leq 2r$

# Application to sparse positive definite forms

Theorem: [Reznick '95] Positivstellensatz

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

# Application to sparse positive definite forms

Theorem: [Reznick '95] Positivstellensatz

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

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

RUNNING INTERSECTION PROPERTY (RIP)

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

# Application to sparse positive definite forms

Theorem: [Reznick '95] Positivstellensatz

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

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

RUNNING INTERSECTION PROPERTY (RIP)

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

Theorem: sparse Reznick [Mai Lasserre Magron '20]

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

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

**Term sparsity**

Ideal sparsity

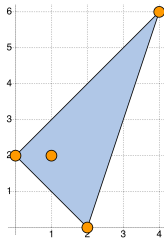
Tutorial session

# Term sparsity via Newton polytope

$$f = 4x_1^4x_2^6 + x_1^2 - x_1x_2^2 + x_2^2$$

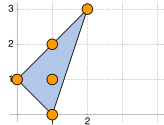
$$\text{spt}(f) = \{(4,6), (2,0), (1,2), (0,2)\}$$

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



$$\text{Squares in SOS decomposition} \subseteq \frac{\mathcal{B}}{2} \cap \mathbb{N}^n$$

[Reznick '78]



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

# Term sparsity: the unconstrained case

---

$$\begin{aligned} f &= x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 \\ &\quad + 6x_3^2 + 18x_2^2x_3 - 54x_2x_3^2 + 142x_2^2x_3^2 \end{aligned}$$

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

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

# Term sparsity: the unconstrained case

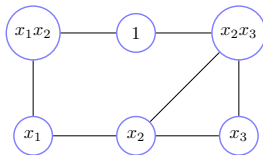
$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 \\ + 6x_3^2 + 18x_2^2x_3 - 54x_2x_3^2 + 142x_2^2x_3^2$$

$$[\text{Reznick '78}] \rightarrow f = (1 \quad x_1 \quad x_2 \quad x_3 \quad x_1x_2 \quad x_2x_3) \underbrace{Q}_{\succeq 0}$$

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

$$\begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$$

💡 **Term sparsity pattern graph**  $G$



# Term sparsity: the unconstrained case

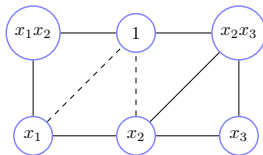
$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 \\ + 6x_3^2 + 18x_2^2x_3 - 54x_2x_3^2 + 142x_2^2x_3^2$$

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

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

$$\begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$$

💡 **Term sparsity pattern graph  $G$**   
+ chordal extension  $G'$



# Term sparsity: the unconstrained case

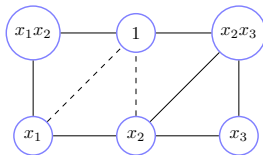
$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 + 6x_3^2 + 18x_2^2x_3 - 54x_2x_3^2 + 142x_2^2x_3^2$$

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

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

$$\begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$$

💡 **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,  $\text{tsp}$  graph  $G$  has

Nodes  $V$  = monomials of degree  $\leq r$

# Term sparsity: the constrained case

---

At step  $r$  of the hierarchy, **tsp** graph  $G$  has

Nodes  $V$  = monomials of degree  $\leq r$

Edges  $E$  with

$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \text{supp } f \bigcup \bigcup_{|\alpha| \leq r} \text{supp } g_j$$

# Term sparsity: the constrained case

At step  $r$  of the hierarchy,  $\text{tsp}$  graph  $G$  has

Nodes  $V$  = monomials of degree  $\leq r$

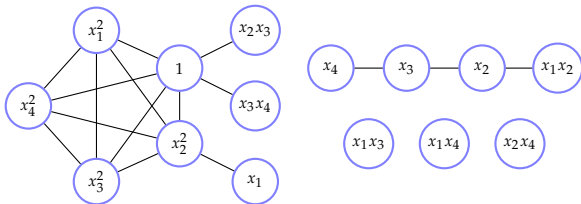
Edges  $E$  with

$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \text{supp } f \bigcup \bigcup_{|\alpha| \leq r} 2\alpha$$

An example with  $r = 2$

$$f = x_1^4 + x_1x_2^2 + x_2x_3 + x_3^2x_4^2$$

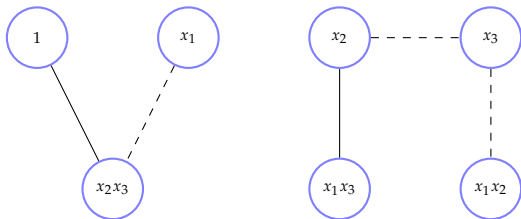
$$g_1 = 1 - x_1^2 - x_2^2 - x_3^2 \quad g_2 = 1 - x_3x_4$$



# Term sparsity: support extension

---

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



## Term sparsity: the constrained case

---

At step  $r$  of the hierarchy,  $\text{tsp}$  graph  $G$  has

Nodes  $V$  = monomials of degree  $\leq r$

# Term sparsity: the constrained case

---

At step  $r$  of the hierarchy, **tsp** graph  $G$  has

Nodes  $V$  = monomials of degree  $\leq r$

Edges  $E$  with

$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \text{supp } f \bigcup \bigcup_{|\alpha| \leq r} \text{supp } g_j$$

$\rightsquigarrow$  **support extension**

# Term sparsity: the constrained case

---

At step  $r$  of the hierarchy, **tsp** graph  $G$  has

Nodes  $V$  = monomials of degree  $\leq r$

Edges  $E$  with

$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \text{supp } f \bigcup \bigcup_{|\alpha| \leq r} \text{supp } g_j$$

$\rightsquigarrow$  **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  $\leq r$

Edges  $E$  with

$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \text{supp } f \bigcup_{|\alpha| \leq r} \text{supp } g_j \bigcup_{|\alpha| \leq r} 2\alpha$$

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

By iteratively performing **support extension** & **chordal extension**

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

💡 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,

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

# 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,

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

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,

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

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

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

# Term sparsity: primal moment relaxations

---

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

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

💡 dual yields the TSSOS hierarchy

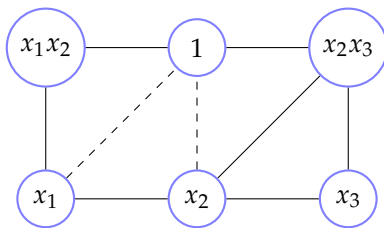
# A two-level hierarchy of lower bounds

---

$$\begin{array}{cccc}
 f_{\text{ts}}^{r_{\min},1} \leq & f_{\text{ts}}^{r_{\min},2} \leq & \dots \leq & f^{r_{\min}} \\
 \wedge | & \wedge | & & \wedge | \\
 f_{\text{ts}}^{r_{\min}+1,1} \leq & f_{\text{ts}}^{r_{\min}+1,2} \leq & \dots \leq & f^{r_{\min}+1} \\
 \wedge | & \wedge | & & \wedge | \\
 \vdots & \vdots & \vdots & \vdots \\
 \wedge | & \wedge | & & \wedge | \\
 f_{\text{ts}}^{r,1} \leq & f_{\text{ts}}^{r,2} \leq & \dots \leq & f^r \\
 \wedge | & \wedge | & & \wedge | \\
 \vdots & \vdots & \vdots & \vdots
 \end{array}$$

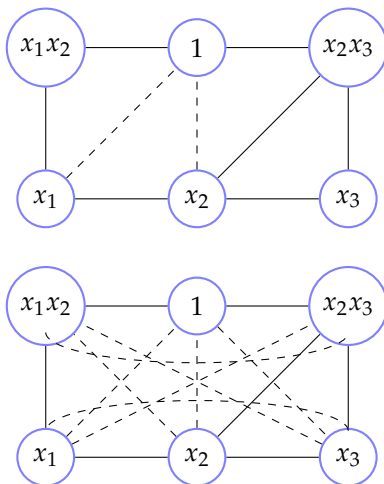
# Different choices of chordal extensions

---



# Different choices of chordal extensions

---



# Term sparsity: convergence guarantees

---

## Theorem [Lasserre Magron Wang '21]

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

# Term sparsity: convergence guarantees

## Theorem [Lasserre Magron Wang '21]

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

💡 Fixing a relaxation order  $r$ , the sequence  $(f_{ts}^{r,s})_{s \geq 1} \uparrow f^r$  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  $(f_{ts}^{r,s})_{r \geq r_{\min}}$  is monotonically non-decreasing.

💡 Fixing a relaxation order  $r$ , the sequence  $(f_{ts}^{r,s})_{s \geq 1} \uparrow f^r$  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** and monomial basis are used.

# Term sparsity: convergence guarantees

## Theorem [Lasserre Magron Wang '21]

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

💡 Fixing a relaxation order  $r$ , the sequence  $(f_{ts}^{r,s})_{s \geq 1} \uparrow f^r$  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** and monomial basis are used.

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

# Term sparsity: convergence guarantees

## Theorem [Lasserre Magron Wang '21]

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

💡 Fixing a relaxation order  $r$ , the sequence  $(f_{ts}^{r,s})_{s \geq 1} \uparrow f^r$  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** and monomial basis are used.

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

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

$$x_2 \mapsto -x_2$$

Sign-symmetries blocks



$$(1 \quad x_1 x_2^2 \quad x_1^2 x_2^2) \quad (x_1 x_2 \quad x_1^2 x_2)$$

**TSSOS** blocks

$$(1 \quad x_1 x_2^2 \quad x_1^2 x_2^2) \quad (x_1 x_2) \quad (x_1^2 x_2)$$

## 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  $\mathbf{Q} \succeq 0 \rightsquigarrow$  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:  $\mathbf{Q}_{ii} \geq \sum_{j \neq i} \mathbf{Q}_{ij} \rightsquigarrow$  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  $\mathbf{Q} \succeq 0 \rightsquigarrow$  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:  $\mathbf{Q}_{ii} \geq \sum_{j \neq i} \mathbf{Q}_{ij} \rightsquigarrow$  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

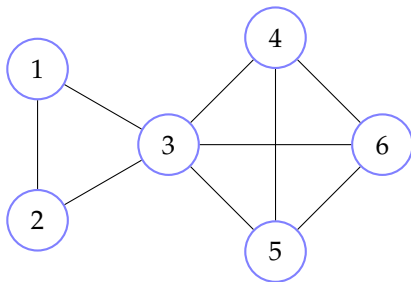
💡 a two-level CS-TSSOS hierarchy of lower bounds for  $f_{\min}$

# 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

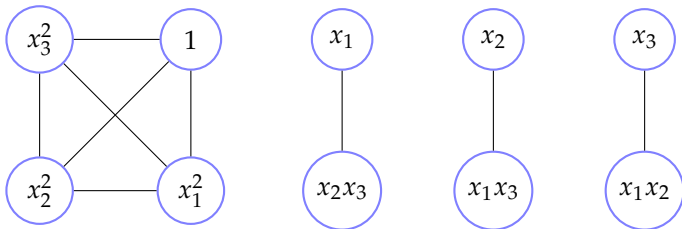


# Combining correlative & term sparsity

---

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

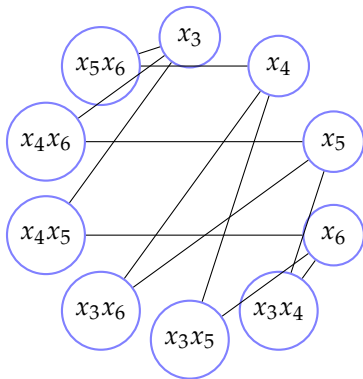
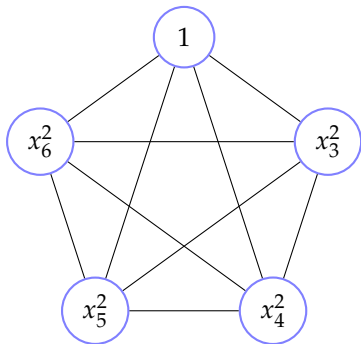
tsp graph for the first clique



# Combining correlative & term sparsity

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

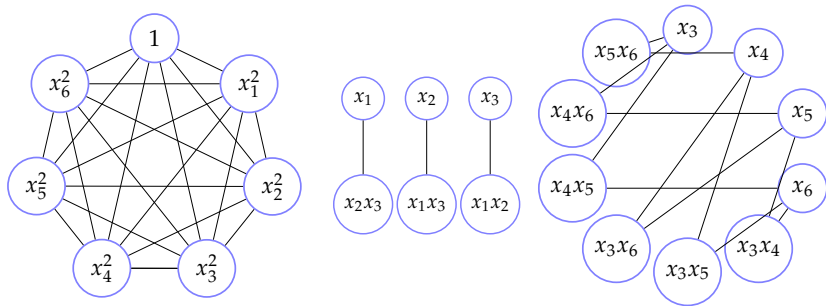
tsp graph for the second clique



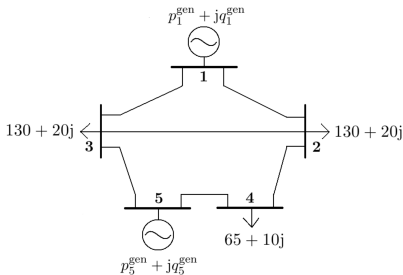
# Combining correlative & term sparsity

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

tsp graph without correlative sparsity



# Application to optimal power-flow



Optimal Powerflow  $n \simeq 10^3$   
[Josz et al. '18]

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

# 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 noncommutative optimization

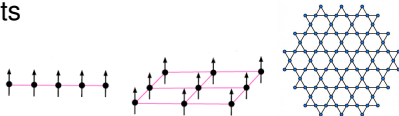
---

Ground-state energy  $\Leftrightarrow$  minimal eigenvalue of an Hamiltonian

$$H = \sum_{\langle i,j \rangle} (x_i x_j + y_i y_j + z_i z_j)$$

spin states  $(x_i, y_i, z_i)$ , constraints

Lattices: 1D 2D Kagome



# Application to noncommutative optimization

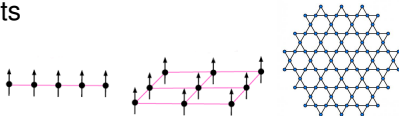
---

Ground-state energy  $\Leftrightarrow$  minimal eigenvalue of an Hamiltonian

$$H = \sum_{\langle i,j \rangle} (x_i x_j + y_i y_j + z_i z_j)$$

spin states  $(x_i, y_i, z_i)$ , constraints

Lattices: 1D 2D Kagome



First neighbors interactions :  $H = \sum_{i=1}^n x_i x_{i+1} + y_i y_{i+1} + z_i z_{i+1}$

# Application to noncommutative optimization

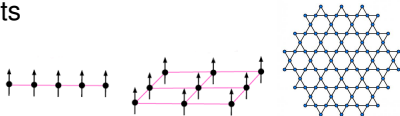
---

Ground-state energy  $\Leftrightarrow$  minimal eigenvalue of an Hamiltonian

$$H = \sum_{\langle i,j \rangle} (x_i x_j + y_i y_j + z_i z_j)$$

spin states  $(x_i, y_i, z_i)$ , constraints

Lattices: 1D 2D Kagome



First neighbors interactions :  $H = \sum_{i=1}^n x_i x_{i+1} + y_i y_{i+1} + z_i z_{i+1}$

periodic boundary conditions  $\Rightarrow$   $\boxed{n+1=1}$

# Application to noncommutative optimization

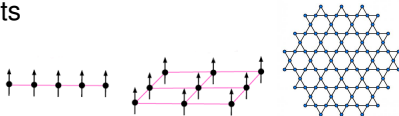
---

Ground-state energy  $\Leftrightarrow$  minimal eigenvalue of an Hamiltonian

$$H = \sum_{\langle i,j \rangle} (x_i x_j + y_i y_j + z_i z_j)$$

spin states  $(x_i, y_i, z_i)$ , constraints

Lattices: 1D 2D Kagome



First neighbors interactions :  $H = \sum_{i=1}^n x_i x_{i+1} + y_i y_{i+1} + z_i z_{i+1}$

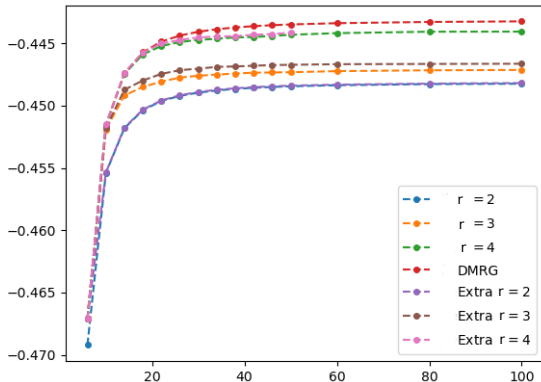
periodic boundary conditions  $\Rightarrow$   $\boxed{n+1=1}$

Existing  $\pm$  efficient techniques: quantum Monte Carlo & variational algorithms  $\Rightarrow$  **upper bounds** on minimal energy

# Application to noncommutative optimization

**Lower bounds** of the energy

1D lattice

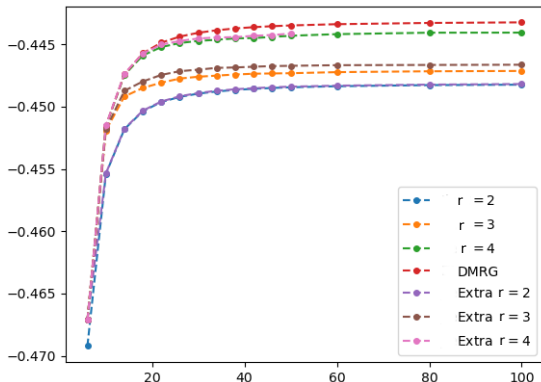


**Dense**  $r = 4, n = 10^2 \Rightarrow 10^{11}$  variables (solvers handle  $\simeq 10^4$ )

# Application to noncommutative optimization

Lower bounds of the energy

1D lattice



**Dense**  $r = 4, n = 10^2 \Rightarrow 10^{11}$  variables (solvers handle  $\simeq 10^4$ )

**Sparse** solved within 1 hour on PFCALCUL at LAAS

# Application to noncommutative optimization

---

## CLASSICAL WORLD

$$\psi^*(A_1 \otimes B_1 + A_1 \otimes B_2 + A_2 \otimes B_1 - A_2 \otimes B_2)\psi \leq 2$$

for separable states  $\psi \in \mathbb{C}^k \otimes \mathbb{C}^k$  and matrices  $A_j, B_j$  satisfying  $A_j^* = A_j, A_j^2 = I, B_j^* = B_j, B_j^2 = I$

# Application to noncommutative optimization

---

## CLASSICAL WORLD

$$\psi^*(A_1 \otimes B_1 + A_1 \otimes B_2 + A_2 \otimes B_1 - A_2 \otimes B_2)\psi \leq 2$$

for separable states  $\psi \in \mathbb{C}^k \otimes \mathbb{C}^k$  and matrices  $A_j, B_j$  satisfying  $A_j^* = A_j, A_j^2 = I, B_j^* = B_j, B_j^2 = I$

**TSIRELSON'S BOUND:**  $2 \rightarrow 2\sqrt{2}$  for maximally entangled states  
 $\psi^*(A \otimes B)\psi = \text{tr}(AB)$

# Application to noncommutative optimization

---

## CLASSICAL WORLD

$$\psi^*(A_1 \otimes B_1 + A_1 \otimes B_2 + A_2 \otimes B_1 - A_2 \otimes B_2)\psi \leq 2$$

for separable states  $\psi \in \mathbb{C}^k \otimes \mathbb{C}^k$  and matrices  $A_j, B_j$  satisfying  $A_j^* = A_j, A_j^2 = I, B_j^* = B_j, B_j^2 = I$

**TSIRELSON'S BOUND:**  $2 \rightarrow 2\sqrt{2}$  for maximally entangled states  
 $\psi^*(A \otimes B)\psi = \text{tr}(AB)$

$$2\sqrt{2} = \text{tr}_{\max}\{a_1b_1 + a_1b_2 + a_2b_1 - a_2b_2 : a_j^2 = b_j^2 = 1\}$$

# Application to noncommutative optimization

---

## COVARIANCES OF QUANTUM CORRELATIONS

$$\text{cov}_\psi(A, B) = \psi^*(A \otimes B)\psi - \psi^*(A \otimes I)\psi \cdot \psi^*(I \otimes B)\psi$$

# Application to noncommutative optimization

---

## COVARIANCES OF QUANTUM CORRELATIONS

$$\text{cov}_\psi(A, B) = \psi^*(A \otimes B)\psi - \psi^*(A \otimes I)\psi \cdot \psi^*(I \otimes B)\psi$$

$$\begin{aligned} & \text{cov}_\psi(A_1, B_1) + \text{cov}_\psi(A_1, B_2) + \text{cov}_\psi(A_1, B_3) \\ & + \text{cov}_\psi(A_2, B_1) + \text{cov}_\psi(A_2, B_2) - \text{cov}_\psi(A_2, B_3) \leq \frac{9}{2} \\ & + \text{cov}_\psi(A_3, B_1) - \text{cov}_\psi(A_3, B_2) \end{aligned}$$

for separable states but ...

# Application to noncommutative optimization

---

## COVARIANCES OF QUANTUM CORRELATIONS

$$\text{cov}_\psi(A, B) = \psi^*(A \otimes B)\psi - \psi^*(A \otimes I)\psi \cdot \psi^*(I \otimes B)\psi$$

$$\begin{aligned} & \text{cov}_\psi(A_1, B_1) + \text{cov}_\psi(A_1, B_2) + \text{cov}_\psi(A_1, B_3) \\ & + \text{cov}_\psi(A_2, B_1) + \text{cov}_\psi(A_2, B_2) - \text{cov}_\psi(A_2, B_3) \leq \frac{9}{2} \\ & + \text{cov}_\psi(A_3, B_1) - \text{cov}_\psi(A_3, B_2) \end{aligned}$$

for separable states but ... 5 for **one** maximally entangled state

# Application to noncommutative optimization

---

## COVARIANCES OF QUANTUM CORRELATIONS

$$\text{cov}_\psi(A, B) = \psi^*(A \otimes B)\psi - \psi^*(A \otimes I)\psi \cdot \psi^*(I \otimes B)\psi$$

$$\begin{aligned} & \text{cov}_\psi(A_1, B_1) + \text{cov}_\psi(A_1, B_2) + \text{cov}_\psi(A_1, B_3) \\ & + \text{cov}_\psi(A_2, B_1) + \text{cov}_\psi(A_2, B_2) - \text{cov}_\psi(A_2, B_3) \leq \frac{9}{2} \\ & + \text{cov}_\psi(A_3, B_1) - \text{cov}_\psi(A_3, B_2) \end{aligned}$$

for separable states but ... 5 for **one** maximally entangled state

💡 2nd dense SDP relaxation of the corresponding trace problem outputs 5

# Application to noncommutative optimization

---

## COVARIANCES OF QUANTUM CORRELATIONS

$$\text{cov}_\psi(A, B) = \psi^*(A \otimes B)\psi - \psi^*(A \otimes I)\psi \cdot \psi^*(I \otimes B)\psi$$

$$\begin{aligned} & \text{cov}_\psi(A_1, B_1) + \text{cov}_\psi(A_1, B_2) + \text{cov}_\psi(A_1, B_3) \\ & + \text{cov}_\psi(A_2, B_1) + \text{cov}_\psi(A_2, B_2) - \text{cov}_\psi(A_2, B_3) \leq \frac{9}{2} \\ & + \text{cov}_\psi(A_3, B_1) - \text{cov}_\psi(A_3, B_2) \end{aligned}$$

for separable states but ... 5 for **one** maximally entangled state

💡 2nd dense SDP relaxation of the corresponding trace problem outputs 5 = max value for **all** maximal entangled states

# Application to noncommutative optimization

---

## COVARIANCES OF QUANTUM CORRELATIONS

$$\text{cov}_\psi(A, B) = \psi^*(A \otimes B)\psi - \psi^*(A \otimes I)\psi \cdot \psi^*(I \otimes B)\psi$$

$$\begin{aligned} & \text{cov}_\psi(A_1, B_1) + \text{cov}_\psi(A_1, B_2) + \text{cov}_\psi(A_1, B_3) \\ & + \text{cov}_\psi(A_2, B_1) + \text{cov}_\psi(A_2, B_2) - \text{cov}_\psi(A_2, B_3) \leq \frac{9}{2} \\ & + \text{cov}_\psi(A_3, B_1) - \text{cov}_\psi(A_3, B_2) \end{aligned}$$

for separable states but ... 5 for **one** maximally entangled state

💡 2nd dense SDP relaxation of the corresponding trace problem outputs 5 = max value for **all** maximal entangled states

💡 2nd sparse SDP gives also 5 ... **10 times faster**

# Application to networked systems stability

---

## Lyapunov function

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

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

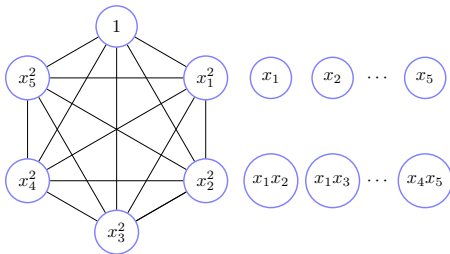
# Application to networked systems stability

## Lyapunov function

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

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

💡 **tsp** graph  $G$



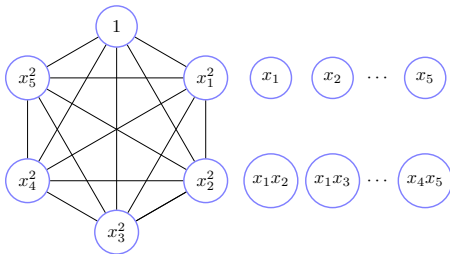
# Application to networked systems stability

## Lyapunov function

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

$\rightsquigarrow \binom{N+2}{2} ((\binom{N+2}{2}) + 1) / 2$  “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  $f \geq 0$  for  $N = 80$  in  $\sim 10$  seconds!

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

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

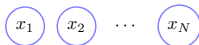
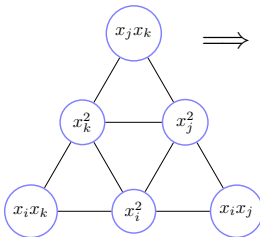
# Application to networked systems stability

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

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

💡 tsp graph  $G$



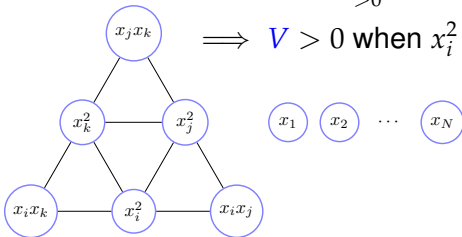
# Application to networked systems stability

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

$\implies V > 0$  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  $f \geq 0$  for  $N = 50$  in  $\sim 1$  second!

# Application to joint spectral radius (JSR)

---

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

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

# Application to joint spectral radius (JSR)

---

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

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

Tons of applications:

- stability of switched linear systems
- continuity of wavelet functions
- trackability of graphs
- ...

# Application to joint spectral radius (JSR)

---

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

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

Tons of applications:

- stability of switched linear systems
- continuity of wavelet functions
- trackability of graphs
- ...

... NP-hard to compute/approximate

# Application to joint spectral radius (JSR)

## Theorem [Parrilo & Jadbabaie '08]

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

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

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

# Application to joint spectral radius (JSR)

## Theorem [Parrilo & Jadbabaie '08]

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

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

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

$$\begin{aligned} \text{💡 } \rho(\mathcal{A}) &\leq \rho^r(\mathcal{A}) = \inf_{f \in \mathbb{R}[\mathbf{x}]_{2r}, \gamma} \gamma \\ \text{s.t. } &\begin{cases} f(\mathbf{x}) - \|\mathbf{x}\|_2^{2r} \text{ SOS} \\ \gamma^{2r} f(\mathbf{x}) - f(A_i \mathbf{x}) \text{ SOS} \end{cases} \end{aligned}$$

# Application to joint spectral radius (JSR)

## Theorem [Parrilo & Jadbabaie '08]

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

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

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

$$\begin{aligned} \text{💡 } \rho(\mathcal{A}) &\leq \rho^r(\mathcal{A}) = \inf_{f \in \mathbb{R}[\mathbf{x}]_{2r}, \gamma} \gamma \\ \text{s.t. } &\begin{cases} f(\mathbf{x}) - \|\mathbf{x}\|_2^{2r} \text{ SOS} \\ \gamma^{2r} f(\mathbf{x}) - f(A_i \mathbf{x}) \text{ SOS} \end{cases} \end{aligned}$$

Bisection on  $\gamma$  + 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”  $f$

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

# Application to joint spectral radius (JSR)

---

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

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

Recursively,  $f_{s-1} = \sum_{\alpha \in \mathcal{A}^{(s-1)}} c_\alpha \mathbf{x}^\alpha$  and

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

# Application to joint spectral radius (JSR)

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

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

Recursively,  $f_{s-1} = \sum_{\alpha \in \mathcal{A}^{(s-1)}} c_\alpha \mathbf{x}^\alpha$  and

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

Theorem: Sparse JSR [Maggio Magron Wang '21]

$$\begin{aligned} \text{💡 } \rho(\mathcal{A}) &\leq \rho^r(\mathcal{A}) \leq \rho^{r,s}(\mathcal{A}) = \inf_{f \in \mathbb{R}[\mathcal{A}^{(s)}], \gamma} \gamma \\ \text{s.t. } &\begin{cases} f(\mathbf{x}) - \|\mathbf{x}\|_2^{2r} \text{ SOS}(\mathcal{A}^{(s)}) \\ \gamma^{2r} f(\mathbf{x}) - f(A_i \mathbf{x}) \text{ SOS}(\mathcal{A}_i^{(s)}) \end{cases} \end{aligned}$$

# 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!

Sparse SDP

Correlative sparsity

Term sparsity

**Ideal sparsity**

Tutorial session

# Ideal sparsity

---

$$f_{\min} = \inf\{f(x_1, x_2) : x_1 x_2 = 0\}$$

# Ideal sparsity

---

$$f_{\min} = \inf\{f(x_1, x_2) : x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, x_2) - \lambda \geq 0 \text{ whenever } x_1 x_2 = 0\}$$

# Ideal sparsity

---

$$f_{\min} = \inf\{f(x_1, x_2) : x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, x_2) - \lambda \geq 0 \text{ whenever } x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, 0) - \lambda \geq 0, \quad f(0, x_2) - \lambda \geq 0\}$$

# Ideal sparsity

---

$$f_{\min} = \inf\{f(x_1, x_2) : x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, x_2) - \lambda \geq 0 \text{ whenever } x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, 0) - \lambda \geq 0, \quad f(0, x_2) - \lambda \geq 0\}$$

💡 replace  $f(x_1, 0) - \lambda \geq 0$  by  $f(x_1, 0) - \lambda = \sigma_1(x_1)$  with SOS  $\sigma_1$

# Ideal sparsity

---

$$f_{\min} = \inf\{f(x_1, x_2) : x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, x_2) - \lambda \geq 0 \text{ whenever } x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, 0) - \lambda \geq 0, \quad f(0, x_2) - \lambda \geq 0\}$$

💡 replace  $f(x_1, 0) - \lambda \geq 0$  by  $f(x_1, 0) - \lambda = \sigma_1(x_1)$  with SOS  $\sigma_1$

General ideal constraints  $x_i x_j = 0 \quad \forall (i, j) \in \overline{E}$

# Ideal sparsity

---

$$f_{\min} = \inf\{f(x_1, x_2) : x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, x_2) - \lambda \geq 0 \text{ whenever } x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, 0) - \lambda \geq 0, \quad f(0, x_2) - \lambda \geq 0\}$$

💡 replace  $f(x_1, 0) - \lambda \geq 0$  by  $f(x_1, 0) - \lambda = \sigma_1(x_1)$  with SOS  $\sigma_1$

General ideal constraints  $x_i x_j = 0 \quad \forall (i, j) \in \overline{E}$

$\rightsquigarrow$  max. cliques of the graph with vertices  $\{1, \dots, n\}$  & edges  $E$

# Ideal sparsity

$$f_{\min} = \inf\{f(x_1, x_2) : x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, x_2) - \lambda \geq 0 \text{ whenever } x_1 x_2 = 0\}$$

$$= \sup\{\lambda : f(x_1, 0) - \lambda \geq 0, \quad f(0, x_2) - \lambda \geq 0\}$$

💡 replace  $f(x_1, 0) - \lambda \geq 0$  by  $f(x_1, 0) - \lambda = \sigma_1(x_1)$  with SOS  $\sigma_1$

General ideal constraints  $x_i x_j = 0 \quad \forall (i, j) \in \overline{E}$

$\rightsquigarrow$  max. cliques of the graph with vertices  $\{1, \dots, n\}$  & edges  $E$

**Theorem [Korda-Laurent-Magron-Steenkamp '22]**

Ideal-sparse hierarchies provide better bounds than the dense ones



ACCURACY

# Application to matrix ranks

---

Given a symmetric nonnegative matrix  $A$ , find the smallest  $r_+$  s.t.

$$A = \sum_{\ell=1}^{r_+} a_\ell a_\ell^T \quad \text{for } a_\ell \geq 0$$

$r_+$  is called the completely positive rank

# Application to matrix ranks

---

Given a symmetric nonnegative matrix  $A$ , find the smallest  $r_+$  s.t.

$$A = \sum_{\ell=1}^{r_+} a_\ell a_\ell^T \quad \text{for } a_\ell \geq 0$$

$r_+$  is called the completely positive rank

✗ hard to compute

# Application to matrix ranks

---

Given a symmetric nonnegative matrix  $A$ , find the smallest  $r_+$  s.t.

$$A = \sum_{\ell=1}^{r_+} a_\ell a_\ell^T \quad \text{for } a_\ell \geq 0$$

$r_+$  is called the completely positive rank

✗ hard to compute

✓ Relax/convexify with a linear program over measures

$$r_+ \geq \inf_{\mu} \left\{ \int_{K_A} 1 d\mu : \int_{K_A} x_i x_j d\mu = A_{ij} \ (i, j \in V), \quad \text{supp}(\mu) \subseteq K_A \right\}$$

# Application to matrix ranks

---

Given a symmetric nonnegative matrix  $A$ , find the smallest  $r_+$  s.t.

$$A = \sum_{\ell=1}^{r_+} a_\ell a_\ell^T \quad \text{for } a_\ell \geq 0$$

$r_+$  is called the completely positive rank

✗ hard to compute

✓ Relax/convexify with a linear program over measures

$$r_+ \geq \inf_{\mu} \left\{ \int_{K_A} 1 d\mu : \int_{K_A} x_i x_j d\mu = A_{ij} \ (i, j \in V), \quad \text{supp}(\mu) \subseteq K_A \right\}$$

$$K_A = \{ \mathbf{x} : \sqrt{A_{ii}} x_i - x_i \geq 0, \quad A_{ij} - x_i x_j \geq 0 \ (i, j) \in E_A, \\ x_i x_j = 0 \ (i, j) \in \bar{E}_A, \quad A - \mathbf{x} \mathbf{x}^T \succcurlyeq 0 \}$$

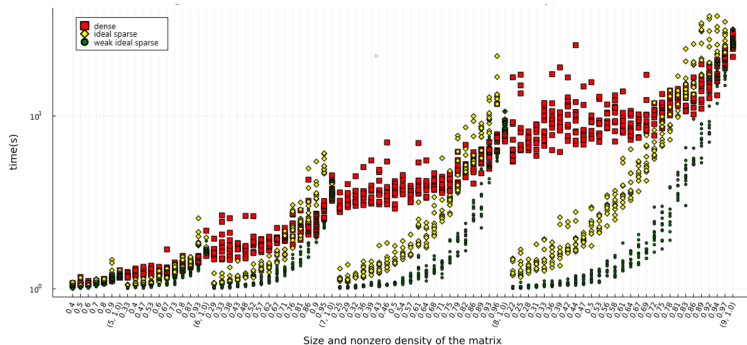
# Application to matrix ranks

---

Random instances, order 2

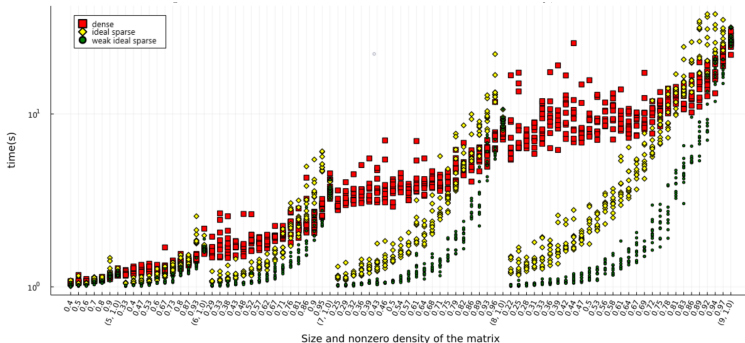
# Application to matrix ranks

## Random instances, order 2



# Application to matrix ranks

Random instances, order 2



PERFORMANCE



AND



ACCURACY

# Conclusion

---

SPARSITY EXPLOITING CONVERGING HIERARCHIES to minimize  
polynomials, eigenvalue/trace, joint spectral radii

# Conclusion

---

SPARSITY EXPLOITING CONVERGING HIERARCHIES to minimize  
polynomials, eigenvalue/trace, joint spectral radii

FAST IMPLEMENTATION IN JULIA: TSSOS, NCTSSOS, SparseJSR

# Conclusion

---

SPARSITY EXPLOITING CONVERGING HIERARCHIES to minimize  
polynomials, eigenvalue/trace, joint spectral radii

FAST IMPLEMENTATION IN JULIA: TSSOS, NCTSSOS, SparseJSR

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

# Conclusion

---

SPARSITY EXPLOITING CONVERGING HIERARCHIES to minimize polynomials, eigenvalue/trace, joint spectral radii

FAST IMPLEMENTATION IN JULIA: TSSOS, NCTSSOS, SparseJSR

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



## Further topics

---

**Correlative** sparsity: Convergence rate?



# Further topics

---

**Correlative** sparsity: Convergence rate?

**Term** sparsity: Smart solution extraction?



## Further topics

---

**Correlative** sparsity: Convergence rate?

**Term** sparsity: Smart solution extraction?

**Ideal** sparsity: tensor (nonnegative, symmetric) ranks?



## Further topics

---

**Correlative** sparsity: Convergence rate?

**Term** sparsity: Smart solution extraction?

**Ideal** sparsity: tensor (nonnegative, symmetric) ranks?

Numerical conditioning of sparse SDP?



## Further topics

---

**Correlative** sparsity: Convergence rate?

**Term** sparsity: Smart solution extraction?

**Ideal** sparsity: tensor (nonnegative, symmetric) ranks?

Numerical conditioning of sparse SDP?

Combination with symmetries?



## Further topics

---

**Correlative** sparsity: Convergence rate?

**Term** sparsity: Smart solution extraction?

**Ideal** sparsity: tensor (nonnegative, symmetric) ranks?

Numerical conditioning of sparse SDP?

Combination with symmetries?

💡 Tons of applications!



# 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

Term sparsity

Ideal sparsity

**Tutorial session**

# Newton polytope

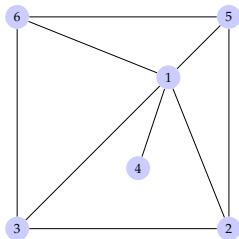
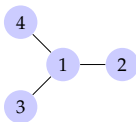
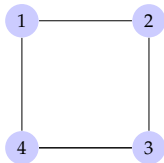
---

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

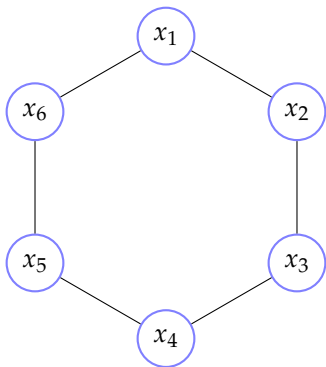
# Chordal or not chordal?

---



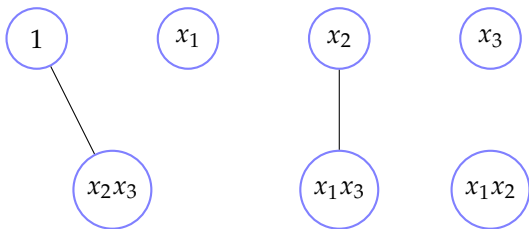
# Chordal extension

---



# Support extension

---

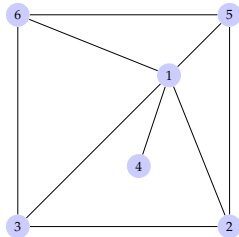


# How big is CS?

---

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

Chordal graph after adding edge (3,5)

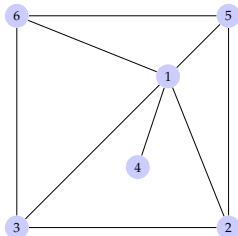


# How big is CS?

---

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

Chordal graph after adding edge (3,5)



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

# Moment matrix

---

Write the first (correlative) sparse moment relaxation of

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

# 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

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

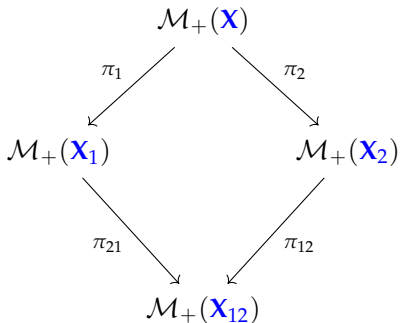
# Measure LP preserves sparsity

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

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

s.t.  $\pi_{12}\mu_1 = \pi_{21}\mu_2$   
 $\mu_1 \in \mathcal{M}_+(\mathbf{X}_1), \quad \mu_2 \in \mathcal{M}_+(\mathbf{X}_2)$

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

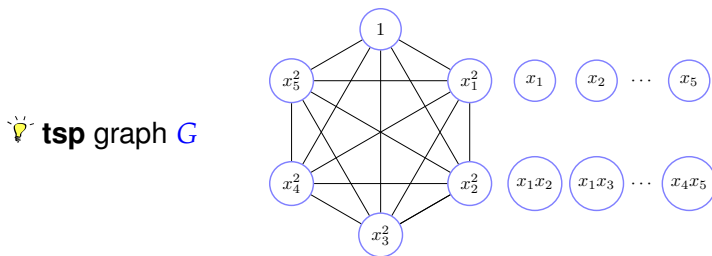


$$f = \sum_{i=1}^N (x_i^2 + x_i^4) - \sum_{i,k=1}^N x_i^2 x_k^2$$

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

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

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



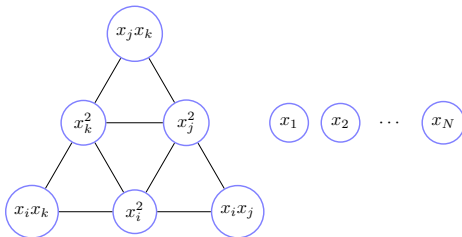
$$f = \sum_{i=1}^N \left( \frac{x_i^2}{2} - \frac{x_i^4}{4} \right) + \sum_{i,k=1}^N (x_i - x_k)^4$$

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

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

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

💡 **tsp** graph  $G$



$$f_1 = x_1^4 + (x_1x_2 - 1)^2 \quad f_2 = x_2^2x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$$

$$f_1 = x_1^4 + (x_1x_2 - 1)^2 \quad f_2 = x_2^2x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$$

Compute the dense relaxation  $f^2$

Compare with the correlative sparse relaxation  $f_{cs}^2$

Compare with the term sparse relaxation  $f_{ts}^{2,s}$  for  $s = 1, 2, \dots$

$$f_1 = x_1^4 + (x_1x_2 - 1)^2 \quad f_2 = x_2^2x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$$

Compute the dense relaxation  $f^2$

Compare with the correlative sparse relaxation  $f_{cs}^2$

Compare with the term sparse relaxation  $f_{ts}^{2,s}$  for  $s = 1, 2, \dots$

💡 Install and run TSSOS:

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

$$f_1 = x_1^4 + (x_1x_2 - 1)^2 \quad f_2 = x_2^2x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$$

Compute the dense relaxation  $f^2$

Compare with the correlative sparse relaxation  $f_{cs}^2$

Compare with the term sparse relaxation  $f_{ts}^{2,s}$  for  $s = 1, 2, \dots$

💡 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
```

$$f_1 = x_1^4 + (x_1x_2 - 1)^2 \quad f_2 = x_2^2x_3^2 + (x_3^2 - 1)^2 \quad f = f_1 + f_2$$

Compute the dense relaxation  $f^2$

Compare with the correlative sparse relaxation  $f_{cs}^2$

Compare with the term sparse relaxation  $f_{ts}^{2,s}$  for  $s = 1, 2, \dots$

💡 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
dense2,sol,data=cs_tssos_first([f], x, 2,
CS=false,TS=false);
```

Download from <https://homepages.laas.fr/vmagron/ncball>:

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

Download from <https://homepages.laas.fr/vmagron/ncball>:

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

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

Download from <https://homepages.laas.fr/vmagron/ncball>:

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

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

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

Download from <https://homepages.laas.fr/vmagron/ncball>:

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

Compute  $\lambda_{\min}(f)$  on  $\mathbb{B}_{nc}$  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