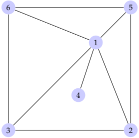
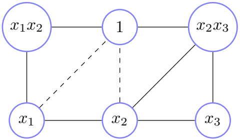
Sparsity in Polynomial Optimization

Victor Magron (LAAS CNRS)

Aromath seminar
13 November 2022



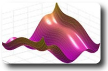




Dense polynomial optimization

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

Theory

(Primal)		(Dual)
$\inf \int f \, d\mu$		$\sup \; b$
with μ proba $\Rightarrow$	**INFINITE-DIM**	$\Leftarrow$ with $f - b \geqslant 0$

Dense polynomial optimization

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

Practice

(Primal **Relaxation**)

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

(Dual **Strengthening**)

$f - b =$ **sum of squares**

finite number $\Rightarrow$ **FINITE-DIM** $\Leftarrow$ **fixed** degree

[Lasserre '01] HIERARCHY of **CONVEX PROBLEMS** $\uparrow f_{\min}$
Based on representing positive polynomials [Putinar '93]

Dense polynomial optimization

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

Practice

(Primal **Relaxation**)

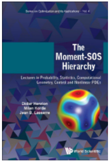
moments $\int \mathbf{x}^\alpha \, d\mu$

(Dual **Strengthening**)

$f - b =$ **sum of squares**

finite number $\Rightarrow$ **FINITE-DIM** $\Leftarrow$ **fixed** degree

[Lasserre '01] HIERARCHY of **CONVEX PROBLEMS** $\uparrow f_{\min}$
Based on representing positive polynomials [Putinar '93]



Attracted a lot of attention in optimization, applied mathematics, quantum computing, engineering, theoretical computer science

Sparse polynomial optimization

Structure exploitation with "SPARSE" cost f and constraints

Sparse polynomial optimization

Structure exploitation with "SPARSE" cost f and constraints

Correlative sparsity: few variable products in f

Sparse polynomial optimization

Structure exploitation with "SPARSE" cost f and constraints

Correlative sparsity: few variable products in f

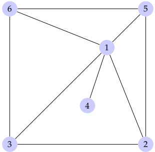
$\rightsquigarrow f = x_1 x_2 + x_2 x_3 + \cdots + x_{99} x_{100}$

Sparse polynomial optimization

Structure exploitation with "SPARSE" cost f and constraints

Correlative sparsity: few variable products in f

$\rightsquigarrow f = x_1 x_2 + x_2 x_3 + \cdots + x_{99} x_{100}$



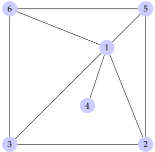
Sparse polynomial optimization

Structure exploitation with "SPARSE" cost f and constraints

Correlative sparsity: few variable products in f

$\rightsquigarrow f = x_1 x_2 + x_2 x_3 + \cdots + x_{99} x_{100}$

Term sparsity: few terms in f



Sparse polynomial optimization

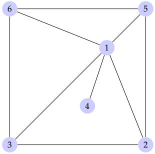
Structure exploitation with "SPARSE" cost f and constraints

Correlative sparsity: few variable products in f

$\rightsquigarrow f = x_1 x_2 + x_2 x_3 + \cdots + x_{99} x_{100}$

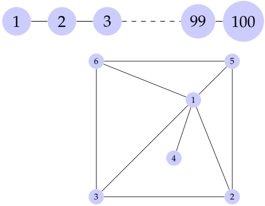
Term sparsity: few terms in f

$\rightsquigarrow f = x_1^{99} x_2 + x_1 x_2^{100}$



Sparse polynomial optimization

Structure exploitation with "SPARSE" cost f and constraints



Correlative sparsity: few variable products in f

$\rightsquigarrow f = x_1 x_2 + x_2 x_3 + \cdots + x_{99} x_{100}$

Term sparsity: few terms in f

$\rightsquigarrow f = x_1^{99} x_2 + x_1 x_2^{100}$

Ideal sparsity: constraints

Sparse polynomial optimization

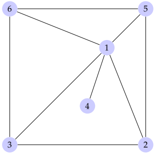
Structure exploitation with "SPARSE" cost f and constraints

Correlative sparsity: few variable products in f

$\rightsquigarrow f = x_1 x_2 + x_2 x_3 + \cdots + x_{99} x_{100}$

Term sparsity: few terms in f

$\rightsquigarrow f = x_1^{99} x_2 + x_1 x_2^{100}$



Ideal sparsity: constraints

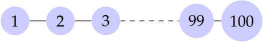
$\rightsquigarrow x_1 x_2 = x_2 x_3 = 0$

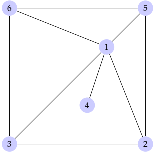
Sparse polynomial optimization

Structure exploitation with "SPARSE" cost f and constraints

Correlative sparsity: few variable products in f

$\rightsquigarrow f = x_1 x_2 + x_2 x_3 + \cdots + x_{99} x_{100}$



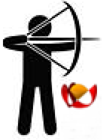


Term sparsity: few terms in f

$\rightsquigarrow f = x_1^{99} x_2 + x_1 x_2^{100}$

Ideal sparsity: constraints

$\rightsquigarrow x_1 x_2 = x_2 x_3 = 0$

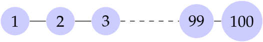
PERFORMANCE  VS  ACCURACY

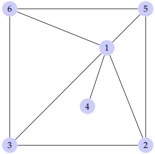
Sparse polynomial optimization

Structure exploitation with "SPARSE" cost f and constraints

Correlative sparsity: few variable products in f

$\rightsquigarrow f = x_1 x_2 + x_2 x_3 + \cdots + x_{99} x_{100}$



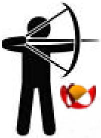


Term sparsity: few terms in f

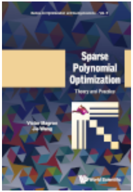
$\rightsquigarrow f = x_1^{99} x_2 + x_1 x_2^{100}$

Ideal sparsity: constraints

$\rightsquigarrow x_1 x_2 = x_2 x_3 = 0$

PERFORMANCE  VS  ACCURACY

Tons of applications: computer arithmetic, deep learning, entanglement, optimal power-flow, analysis of dynamical systems, matrix ranks



Where do we find sparse POPs?

Where do we find sparse POPs?

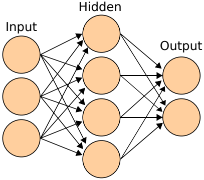
Everywhere!

Where do we find sparse POPs?

Everywhere!

Deep learning

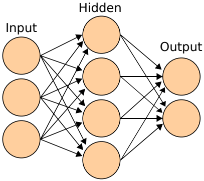
$\rightsquigarrow$ robustness, computer vision



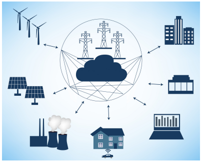
Where do we find sparse POPs?

Everywhere!

Deep learning
⇝ robustness, computer vision



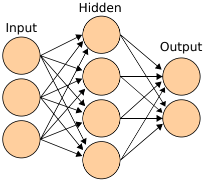
Power systems
⇝ AC optimal power-flow, stability



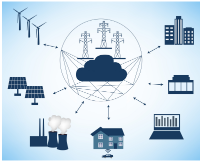
Where do we find sparse POPs?

Everywhere!

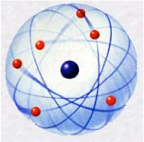
Deep learning
$\rightsquigarrow$ robustness, computer vision



Power systems
$\rightsquigarrow$ AC optimal power-flow, stability



Quantum Systems
$\rightsquigarrow$ condensed matter, entanglement



Moment-SOS hierarchies

Correlative sparsity

Term sparsity

Ideal sparsity

Moment-SOS hierarchies

Correlative sparsity

Term sparsity

Ideal sparsity

Moment-SOS hierarchies: an example

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

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

Moment-SOS hierarchies: an example

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

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

Moment-SOS hierarchies: an example

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

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

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

Moment-SOS hierarchies: an example

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

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

$$\overbrace{x_1 x_2}^{f} =$$

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

Sums of squares (SOS) σ_j

Moment-SOS hierarchies: an example

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

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

$$\overbrace{x_1 x_2}^{f} =$$

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

Sums of squares (SOS) σ_j

Quadratic module: $\mathcal{M}(\mathbf{X})_d = \left\{ \sigma_0 + \sum_j \sigma_j g_j, \ \deg \sigma_j \, g_j \leqslant 2d \right\}$

Moment-SOS hierarchies

Hierarchy of SDP relaxations:

$$\lambda_d := \sup_{\lambda} \Big\{ \lambda : f - \lambda \in \mathcal{M}(\mathbf{X})_d \Big\}$$

Moment-SOS hierarchies

Hierarchy of SDP relaxations:

$$\lambda_d := \sup_{\lambda} \left\{ \lambda : f - \lambda \in \mathcal{M}(\mathbf{X})_d \right\}$$

✓ Convergence guarantees $\lambda_d \uparrow f_{\min}$ [Lasserre '01] when $N - \sum x_i^2 \in \mathcal{M}(\mathbf{X})$ for some $N > 0$

Moment-SOS hierarchies

Hierarchy of SDP relaxations:

$$\lambda_d := \sup_{\lambda} \left\{ \lambda : f - \lambda \in \mathcal{M}(\mathbf{X})_d \right\}$$

✓ Convergence guarantees $\lambda_d \uparrow f_{\min}$ [Lasserre '01] when $N - \sum x_i^2 \in \mathcal{M}(\mathbf{X})$ for some $N > 0$

✓ Can be computed with SDP solvers (`CSDP`, `SDPA`, `MOSEK`)

Moment-SOS hierarchies

Hierarchy of SDP relaxations:

$$\lambda_d := \sup_{\lambda} \left\{ \lambda : f - \lambda \in \mathcal{M}(\mathbf{X})_d \right\}$$

✓ Convergence guarantees $\lambda_d \uparrow f_{\min}$ [Lasserre '01] when $N - \sum x_i^2 \in \mathcal{M}(\mathbf{X})$ for some $N > 0$

✓ Can be computed with SDP solvers (`CSDP`, `SDPA`, `MOSEK`)

✗ **"No Free Lunch" Rule**: $\binom{n+2d}{n}$ SDP variables

Moment-SOS hierarchies

Correlative sparsity

Term sparsity

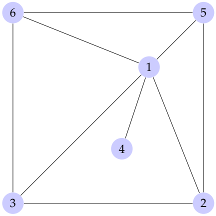
Ideal sparsity

Correlative sparsity

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

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

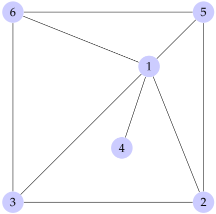
Chordal graph after adding edge $(3, 5)$



Correlative sparsity

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

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



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

maximal cliques I_k

Average size $\kappa \rightsquigarrow \kappa^{2d}$ vars

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

Dense SDP: 210 vars

Sparse SDP: 115 vars

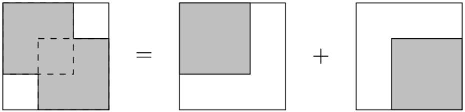
Correlative sparsity

Theorem [Griewank Toint '84]

Chordal graph G with maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries at edges of G

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



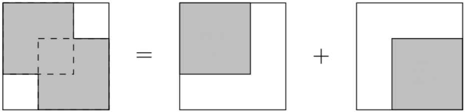
Correlative sparsity

Theorem [Griewank Toint '84]

Chordal graph G with maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries at edges of G

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



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

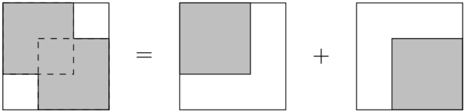
Correlative sparsity

Theorem [Griewank Toint '84]

Chordal graph G with maximal cliques I_1, I_2

$Q_G \succcurlyeq 0$ with nonzero entries at edges of G

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



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

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

$f > 0$ on $\{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$
chordal graph G with cliques I_k $\Longrightarrow$
ball constraints for each $\mathbf{x}(I_k)$

$$\boxed{f = \sigma_{01} + \sigma_{02} + \sum_j \sigma_j g_j}$$

SOS σ_{0k} "sees" vars in I_k
σ_j "sees" vars from g_j

Application to robustness of neural networks

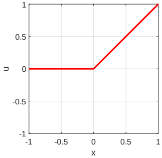
[SIAM News March '21]

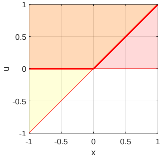
"Yet DL has an Achilles' heel. Current implementations can be highly unstable, meaning that a certain small perturbation to the input of a trained neural network can cause substantial change in its output. This phenomenon is both a nuisance and a major concern for the safety and robustness of DL-based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

Application to robustness of neural networks

$$\mathbf{x}_0 \in \mathbb{R}^p \longrightarrow \mathbf{z}_0 \in \mathbb{R}^p \longrightarrow \mathbf{z}_1 \in \mathbb{R}^{p_1} \longrightarrow \cdots \longrightarrow \mathbf{z}_m \in \mathbb{R}^{p_m} \quad \mathbf{z}_i = \mathbf{A}_i \mathrm{ReLU}(\mathbf{z}_{i-1}) + \mathbf{b}_i$$

ReLU (left) & its "semialgebraicity" (right)





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

Application to robustness of neural networks

"Direct" certification of a classifier with 1 hidden layer

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

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

Application to robustness of neural networks

“Direct” certification of a classifier with 1 hidden layer

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

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

Monotone equilibrium networks [Winston Kolter ’20]

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

Application to robustness of neural networks

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

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

- Monotone equilibrium networks [Winston Kolter '20]

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

- "Indirect" with Lipschitz constant/ellipsoid approximation

Application to robustness of neural networks

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

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

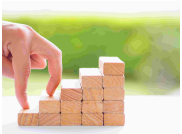
$$\text{s.t.} \quad \begin{cases} \mathbf{z} = \mathrm{ReLU}(\mathbf{A}\mathbf{x} + \mathbf{b}) \\ ||\mathbf{x} - \mathbf{x}_0|| \leq \epsilon \end{cases}$$

- Monotone equilibrium networks [Winston Kolter '20]

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

- "Indirect" with Lipschitz constant/ellipsoid approximation

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



Trained $(784, 500)$ network

MNIST classifier [Raghunathan et al. '18]

Trained $(784, 500)$ network

MNIST classifier [Raghunathan et al. ’18]

Shor: relaxation given by 1ST stair in the hierarchy
LipOpt: LP based method
Sampling: lower bound given by 10^4 random samples

Trained $(784, 500)$ network

MNIST classifier [Raghunathan et al. '18]

Shor: relaxation given by 1ST stair in the hierarchy
LipOpt: LP based method
Sampling: lower bound given by 10^4 random samples

	Sparse		Shor	LipOpt	Sampling
Bound	14.56	$<$	17.85	Out of RAM	9.69
Time	12246	$>$	2869	Out of RAM	-

Trained $(784, 500)$ network

MNIST classifier [Raghunathan et al. '18]

Shor: relaxation given by 1ST stair in the hierarchy
LipOpt: LP based method
Sampling: lower bound given by 10^4 random samples

	Sparse		**Shor**	**LipOpt**	**Sampling**
Bound	14.56	$<$	17.85	Out of RAM	9.69
Time	12246	$>$	2869	Out of RAM	-

PERFORMANCE  VS  ACCURACY

Extension to noncommutative optimization

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

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 $x = (x_1, \ldots, x_n)$

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 $\not\Rightarrow f$ is a sparse SOS

Extension to noncommutative optimization

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

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 $\not\Rightarrow f$ is a sparse SOS

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

Extension to noncommutative optimization

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

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 $\not\Rightarrow f$ is a sparse SOS 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 $x = (x_1, \ldots, x_n)$

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 $\not\Rightarrow f$ is a sparse SOS 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! [Klep Magron Povh '21]

sparse f SOS $\not\Rightarrow f$ is a sparse SOS 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 $\{\mathbf{x} : g_j(\mathbf{x}) \geqslant 0\}$
chordal graph with cliques I_k
ball constraints for each $\mathbf{x}(I_k)$

$\Longrightarrow$

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

s_{ki} "sees" vars in I_k
t_{ji} "sees" vars from g_j

Application to violation of Bell inequalities

$\mathbf{I}_{3322}$ 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 → **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

$\mathbf{I}_{3322}$ 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 $\to$ **upper bounds** on $\lambda_{\max}$ of f on
$\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$

💡 $I_k \to \{a_k, b_1, b_2, b_3\}$

Application to violation of Bell inequalities

$\mathbf{I}_{3322}$ 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 → **upper bounds** on $\lambda_{\max}$ of f on
$\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$

💡 $I_k \to \{a_k, b_1, b_2, b_3\}$

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

Application to violation of Bell inequalities

$\mathbf{I}_{3322}$ 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 → **upper bounds** on $\lambda_{\max}$ of f on
$\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$

💡 $I_k \to \{a_k, b_1, b_2, b_3\}$

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

Application to violation of Bell inequalities

$\mathbf{I}_{3322}$ 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 → **upper bounds** on $\lambda_{\max}$ of f on $\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$

💡 $I_k \to \{a_k, b_1, b_2, b_3\}$

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

$\mathbf{I}_{3322}$ 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 → **upper bounds** on $\lambda_{\max}$ of f on $\{a, b : a_i^2 = a_i \quad b_i^2 = b_i \quad a_i b_j = b_j a_i\}$

💡 $I_k \to \{a_k, b_1, b_2, b_3\}$

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

$\mathbf{I}_{3322}$ 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'		0.250875**4** (1 day)
4	0.2508917	
5	0.2508763	

Application to violation of Bell inequalities

$\mathbf{I}_{3322}$ 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 → **upper bounds** on $\lambda_{\max}$ of f on

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

$I_k \to \{a_k, b_1, b_2, b_3\}$

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	0.25087539771 80	(1 hour)

PERFORMANCE  VS  ACCURACY

Application to violation of Bell inequalities

CLASSICAL WORLD

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

for separable states $\psi \in \mathbb{R}^k \otimes \mathbb{R}^k$ and self-adjoint matrices A_j, B_j satisfying $A_i^2 = B_j^2 = I$

Application to violation of Bell inequalities

CLASSICAL WORLD

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

for separable states $\psi \in \mathbb{R}^k \otimes \mathbb{R}^k$ and self-adjoint matrices A_j, B_j satisfying $A_i^2 = B_j^2 = I$

TSIRELSON'S BOUND for maximally entangled states
$\psi = \frac{1}{\sqrt{k}} \sum_{j=1}^k e_j \otimes e_j \in \mathbb{R}^k \otimes \mathbb{R}^k$

Application to violation of Bell inequalities

CLASSICAL WORLD

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

for separable states $\psi \in \mathbb{R}^k \otimes \mathbb{R}^k$ and self-adjoint matrices A_j, B_j satisfying $A_i^2 = B_j^2 = I$

TSIRELSON'S BOUND for maximally entangled states
$\psi = \frac{1}{\sqrt{k}} \sum_{j=1}^k e_j \otimes e_j \in \mathbb{R}^k \otimes \mathbb{R}^k \implies \psi^*(X \otimes Y)\psi = \mathrm{tr}(XY)$

Application to violation of Bell inequalities

CLASSICAL WORLD

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

for separable states $\psi \in \mathbb{R}^k \otimes \mathbb{R}^k$ and self-adjoint matrices A_j, B_j satisfying $A_i^2 = B_j^2 = I$

TSIRELSON'S BOUND for maximally entangled states
$\psi = \frac{1}{\sqrt{k}} \sum_{j=1}^k e_j \otimes e_j \in \mathbb{R}^k \otimes \mathbb{R}^k \implies \psi^*(X \otimes Y)\psi = \operatorname{tr}(XY)$

$$2 \to 2\sqrt{2} = \operatorname{tr}_{\max}\{a_1 b_1 + a_1 b_2 + a_2 b_1 - a_2 b_2 : a_j^2 = b_j^2 = 1\}$$

Application to violation of Bell inequalities

COVARIANCES OF QUANTUM CORRELATIONS

$$\mathrm{cov}_{\psi}(X, Y) := \psi^*(X \otimes Y)\psi - \psi^*(X \otimes I)\psi \cdot \psi^*(I \otimes Y)\psi$$

Application to violation of Bell inequalities

COVARIANCES OF QUANTUM CORRELATIONS

$$\mathrm{cov}_\psi(X,Y) := \psi^*(X \otimes Y)\psi - \psi^*(X \otimes I)\psi \cdot \psi^*(I \otimes Y)\psi$$

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

for separable states but ...

Application to violation of Bell inequalities

COVARIANCES OF QUANTUM CORRELATIONS

$$\mathrm{cov}_\psi(X,Y) := \psi^*(X \otimes Y)\psi - \psi^*(X \otimes I)\psi \cdot \psi^*(I \otimes Y)\psi$$

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

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

Application to violation of Bell inequalities

COVARIANCES OF QUANTUM CORRELATIONS

$$\mathrm{cov}_\psi(X, Y) := \psi^*(X \otimes Y)\psi - \psi^*(X \otimes I)\psi \cdot \psi^*(I \otimes Y)\psi$$

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

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

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

Application to violation of Bell inequalities

COVARIANCES OF QUANTUM CORRELATIONS

$$\mathrm{cov}_\psi(X, Y) := \psi^*(X \otimes Y)\psi - \psi^*(X \otimes I)\psi \cdot \psi^*(I \otimes Y)\psi$$

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

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 violation of Bell inequalities

COVARIANCES OF QUANTUM CORRELATIONS

$$\mathrm{cov}_\psi(X,Y) := \psi^*(X \otimes Y)\psi - \psi^*(X \otimes I)\psi \cdot \psi^*(I \otimes Y)\psi$$

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

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 5 too . . . 10 times faster

Moment-SOS hierarchies

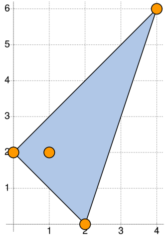
Correlative sparsity

Term sparsity

Ideal sparsity

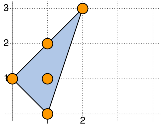
Term sparsity: unconstrained

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



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

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



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

Term sparsity: unconstrained

[Postdoc Wang '19-21] ANR Tremplin-ERC

COPS

$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 + 6x_3^2 + 9x_2^2x_3 - 45x_2x_3^2 + 142x_2^2x_3^2$$

[Reznick '78] → Newton polytope method

$$f = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_2x_1 & x_3x_2 \end{pmatrix} \underbrace{Q}_{\succcurlyeq 0} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$$

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

Term sparsity: unconstrained

[Postdoc Wang '19-21] ANR Tremplin-ERC

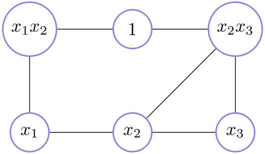
$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 + 6x_3^2 + 9x_2^2x_3 - 45x_2x_3^2 + 142x_2^2x_3^2$$

[Reznick '78] $\to$ Newton polytope method

$$f = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_2x_1 & x_3x_2 \end{pmatrix} \underbrace{Q}_{\succcurlyeq 0} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$$

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

💡 **Term** sparsity pattern graph G



Term sparsity: unconstrained

[Postdoc Wang '19-21] ANR Tremplin-ERC

COPS

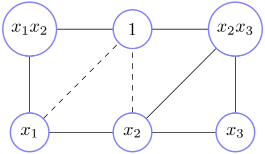
$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 + 6x_3^2 + 9x_2^2x_3 - 45x_2x_3^2 + 142x_2^2x_3^2$$

[Reznick '78] $\rightarrow$ Newton polytope method

$$f = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_2x_1 & x_3x_2 \end{pmatrix} \underbrace{Q}_{\succcurlyeq 0} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$$

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

Term sparsity pattern graph G
+ chordal extension G'



Term sparsity: unconstrained

[Postdoc Wang '19-21] ANR Tremplin-ERC

COPS

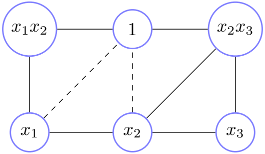
$$f = x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1^2x_2 + 2x_1^2x_2^2 - 2x_2x_3 + 6x_3^2 + 9x_2^2x_3 - 45x_2x_3^2 + 142x_2^2x_3^2$$

[Reznick '78] $\to$ Newton polytope method

$$f = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_2x_1 & x_3x_2 \end{pmatrix} \underbrace{Q}_{\succcurlyeq 0} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_1x_2 \\ x_2x_3 \end{pmatrix}$$

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

Term sparsity pattern graph G
+ chordal extension G'



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

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

Term sparsity: constrained

At step d of the hierarchy, tsp graph G has

Nodes V = monomials of degree $\leqslant d$

Term sparsity: constrained

At step d of the hierarchy, tsp graph G has

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

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

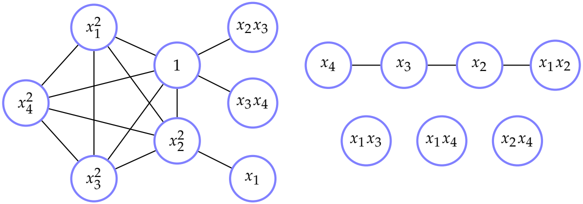
Term sparsity: constrained

At step d of the hierarchy, tsp graph G has

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

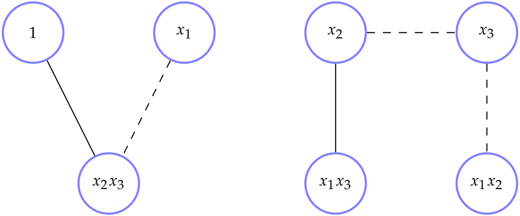
$$\{\alpha, \beta\} \in E \Leftrightarrow \alpha + \beta \in \operatorname{supp} f \bigcup \operatorname{supp} g_j \bigcup_{|\alpha| \leqslant d} 2\alpha$$

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



Term sparsity: support extension

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



Term sparsity: constrained

At step d of the hierarchy, tsp graph G has

Nodes V = monomials of degree $\leqslant d$

Term sparsity: constrained

At step d of the hierarchy, tsp graph G has

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

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

$\rightsquigarrow$ support extension

Term sparsity: constrained

At step d of the hierarchy, tsp graph G has

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

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

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

Term sparsity: constrained

At step d of the hierarchy, tsp graph G has

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

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

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

By iteratively performing support extension & chordal extension

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

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

Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

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

Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

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

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

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

Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

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

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

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

$x_2 \mapsto -x_2$

Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

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

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

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

$x_2 \mapsto -x_2$

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

Term sparsity: convergence guarantees

Theorem [Lasserre Magron Wang '21]

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

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

Newton polytope $\rightsquigarrow \mathscr{B} = (1 \quad x_1x_2 \quad x_1x_2^2 \quad x_1^2x_2 \quad x_1^2x_2^2)$

$x_2 \mapsto -x_2$

Sign symmetries blocks $(1 \quad x_1x_2^2 \quad x_1^2x_2^2) \qquad (x_1x_2 \quad x_1^2x_2)$

Term sparsity blocks $(1 \quad x_1x_2^2 \quad x_1^2x_2^2) \quad (x_1x_2) \quad (x_1^2x_2)$

Term sparsity

CONVERGENCE GUARANTEES

Term sparsity

- CONVERGENCE GUARANTEES
- handles Combo with correlative sparsity

Term sparsity

- CONVERGENCE GUARANTEES

- handles Combo with correlative sparsity

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

Term sparsity

- CONVERGENCE GUARANTEES
- handles Combo with correlative 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 the iterative procedure to exploit term sparsity

Term sparsity

CONVERGENCE GUARANTEES

handles Combo with correlative 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 the iterative procedure to exploit term sparsity

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

Term sparsity

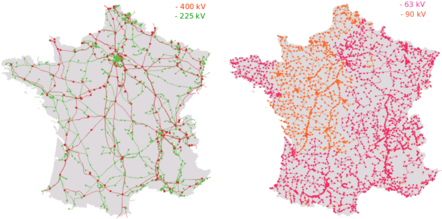
- CONVERGENCE GUARANTEES
- handles Combo with correlative sparsity
  1. Partition the variables w.r.t. the maximal cliques of the csp graph
  2. For each subsystem involving variables from one maximal clique, apply the iterative procedure to exploit term sparsity
- two-level hierarchy of lower bounds for $f_{\min}$: CS-TSSOS hierarchy
- Julia library `TSSOS` $\to$ solve problems with $n = 10^3$

Term sparsity

- CONVERGENCE GUARANTEES
- handles Combo with correlative 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 the iterative procedure to exploit term sparsity
- two-level hierarchy of lower bounds for $f_{\min}$: CS-TSSOS hierarchy
- Julia library `TSSOS` $\rightarrow$ solve problems with $n = 10^3$
- choice of the CHORDAL EXTENSION: min / max

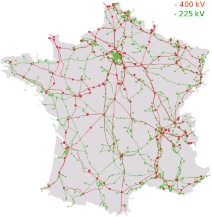
Application to AC optimal power-flow

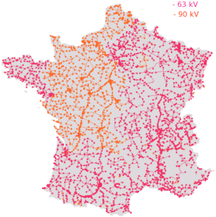
Minimize active power injections of an alternating current transmission network under physical + operational constraints



Application to AC optimal power-flow

Minimize active power injections of an alternating current transmission network under physical + operational constraints





Artificial version of the control problem for electricity transmission network

Application to AC optimal power-flow

Network = Graph with buses N, *from* edges E, *to* edges E^R

Application to AC optimal power-flow

Network = Graph with buses N, *from* edges E, *to* edges E^R

Generators at bus i = G_i, with power demand $\mathbf{S}_i^d$

V_i and S_k^g = voltage at bus i and power generation at generator k

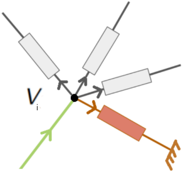
Application to AC optimal power-flow

Network = Graph with buses N, *from* edges E, *to* edges E^R

Generators at bus i = G_i, with power demand $\mathbf{S}_i^d$

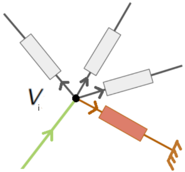
V_i and S_k^g = voltage at bus i and power generation at generator k

Kirchhoff law: $I_i = \sum_{(i,j) \in E_i \cup E_i^R} I_{ij} + I_i^{\text{gr}}$



Application to AC optimal power-flow

Network = Graph with buses N, *from* edges E, *to* edges E^R
Generators at bus i = G_i, with power demand $\mathbf{S}_i^d$
V_i and S_k^g = voltage at bus i and power generation at generator k

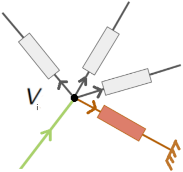


Kirchhoff law: $I_i = \sum_{(i,j) \in E_i \cup E_i^R} I_{ij} + I_i^{\text{gr}}$

Ohm law: $\begin{pmatrix} I_{ij} \\ I_{ji} \end{pmatrix} = \mathbf{Y}_{ij} \begin{pmatrix} V_i \\ V_j \end{pmatrix}$

Application to AC optimal power-flow

Network = Graph with buses N, *from* edges E, *to* edges E^R
Generators at bus i = G_i, with power demand $\mathbf{S}_i^d$
V_i and S_k^g = voltage at bus i and power generation at generator k



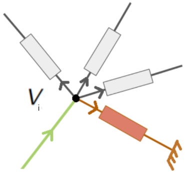
Kirchhoff law: $I_i = \sum_{(i,j) \in E_i \cup E_i^R} I_{ij} + I_i^{\text{gr}}$

Ohm law: $\begin{pmatrix} I_{ij} \\ I_{ji} \end{pmatrix} = \mathbf{Y}_{ij} \begin{pmatrix} V_i \\ V_j \end{pmatrix}$

Relation power-voltage-current: $\sum_{k \in G_i} S_k^g - \mathbf{S}_i^d = V_i I_i^\star$

Application to AC optimal power-flow

Network = Graph with buses N, *from* edges E, *to* edges E^R
Generators at bus i = G_i, with power demand $\mathbf{S}_i^d$
V_i and S_k^g = voltage at bus i and power generation at generator k



Kirchhoff law: $I_i = \sum_{(i,j) \in E_i \cup E_i^R} I_{ij} + I_i^{\text{gr}}$

Ohm law: $\begin{pmatrix} I_{ij} \\ I_{ji} \end{pmatrix} = \mathbf{Y}_{ij} \begin{pmatrix} V_i \\ V_j \end{pmatrix}$

Relation power-voltage-current: $\sum_{k \in G_i} S_k^g - \mathbf{S}_i^d = V_i I_i^\star$
$\rightsquigarrow$ leads to power-flow equations

Application to AC optimal power-flow

mb: maximal block size

gap: the optimality gap w.r.t. local optimal solution

n	m	CS ($d = 2$)			CS-TSSOS ($d = 2, \ell = 1$)		
		mb	time	gap	mb	time	gap
1112	4613	231	3114	0.85%	39	46.6	0.86%
		496	–	–	31	410	0.25%
4356	18257	378	–	–	27	934	0.51%
6698	29283	1326	–	–	76	1886	0.47%

Moment-SOS hierarchies

Correlative sparsity

Term sparsity

Ideal sparsity

Ideal sparsity

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

Ideal sparsity

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

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

Ideal sparsity

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

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

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

Ideal sparsity

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

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

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

replace $f(x_1, 0) - \lambda \geqslant 0$ by $f(x_1, 0) - \lambda = \sigma_1(x_1)$ with SOS σ_1

Ideal sparsity

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

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

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

replace $f(x_1, 0) - \lambda \geqslant 0$ by $f(x_1, 0) - \lambda = \sigma_1(x_1)$ with SOS σ_1

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

Ideal sparsity

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

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

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

💡 replace $f(x_1, 0) - \lambda \geqslant 0$ by $f(x_1, 0) - \lambda = \sigma_1(x_1)$ with SOS σ_1

General ideal constraints $x_i x_j = 0 \quad \forall (i, j) \in \overline{E}$
$\rightsquigarrow$ max. cliques of the graph with vertices $\{1, \ldots, n\}$ & edges E

Ideal sparsity

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

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

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

replace $f(x_1, 0) - \lambda \geqslant 0$ by $f(x_1, 0) - \lambda = \sigma_1(x_1)$ with SOS σ_1

General ideal constraints $x_i x_j = 0 \quad \forall (i, j) \in \overline{E}$
$\rightsquigarrow$ max. cliques of the graph with vertices $\{1, \ldots, n\}$ & 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.

$$A = \sum_{\ell=1}^{r} a_\ell {a_\ell}^T \qquad \text{for } a_\ell \geqslant 0$$

r is called the completely positive rank

Application to matrix ranks

Given a symmetric nonnegative matrix A, find the smallest r s.t.

$$A = \sum_{\ell=1}^{r} a_\ell {a_\ell}^T \qquad \text{for } a_\ell \geqslant 0$$

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.

$$A = \sum_{\ell=1}^{r} a_\ell {a_\ell}^T \qquad \text{for } a_\ell \geqslant 0$$

r is called the completely positive rank
✗ hard to compute
✓ Relax/convexify with a linear program over measures

$$r \geqslant \inf_{\mu} \{ \int_{K_A} 1 d\mu : \int_{K_A} x_i x_j d\mu = A_{ij} \; (i,j \in V) \, , \quad \text{supp}(\mu) \subseteq K_A \}$$

Application to matrix ranks

Given a symmetric nonnegative matrix A, find the smallest r s.t.

$$A = \sum_{\ell=1}^{r} a_\ell {a_\ell}^T \qquad \text{for } a_\ell \geqslant 0$$

r is called the completely positive rank
✗ hard to compute
✓ Relax/convexify with a linear program over measures

$$r \geqslant \inf_{\mu} \{ \int_{K_A} 1 d\mu : \int_{K_A} x_i x_j d\mu = A_{ij} \ (i,j \in V) \,, \quad \text{supp}(\mu) \subseteq K_A \}$$

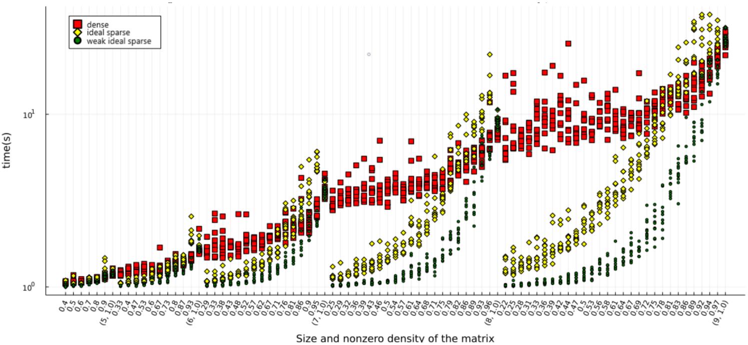
$$K_A = \{ \mathbf{x} : \sqrt{A_{ii}} x_i - x_i \geqslant 0 \,, \quad A_{ij} - x_i x_j \geqslant 0 \ (i,j) \in E_A \,, \\ x_i x_j = 0 \ (i,j) \in \overline{E}_A \,, \quad A - \mathbf{x}\mathbf{x}^T \succcurlyeq 0 \}$$

Application to matrix ranks

Random instances, order 2

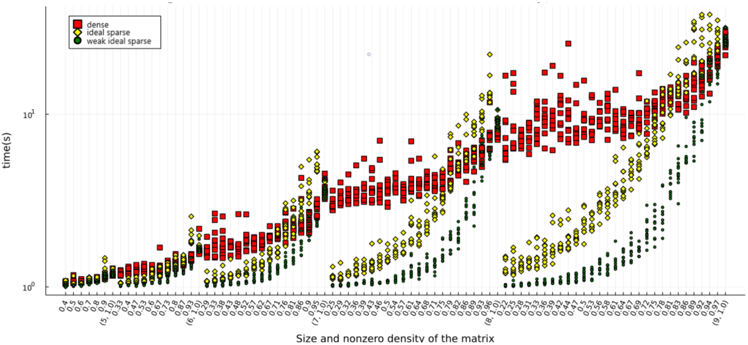
Application to matrix ranks

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 radius

Conclusion

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

FAST IMPLEMENTATION IN JULIA: TSSOS, NCTSSOS, SparseJSR

Conclusion

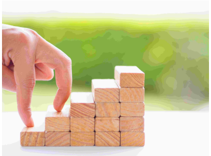
SPARSITY EXPLOITING CONVERGING HIERARCHIES to minimize polynomials, eigenvalue/trace, joint spectral radius

FAST IMPLEMENTATION IN JULIA: TSSOS, NCTSSOS, SparseJSR

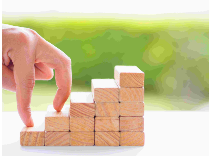
Combine correlative & term sparsity for problems with $n = 10^3$

Further topics

Correlative sparsity: convergence rate?



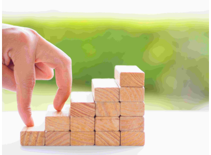
Further topics



Correlative sparsity: convergence rate?

Term sparsity: (smart) solution extraction

Further topics

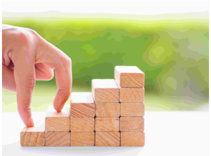


Correlative sparsity: convergence rate?

Term sparsity: (smart) solution extraction

Ideal sparsity: tensor ranks?

Further topics



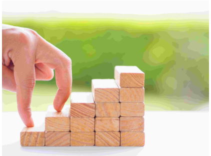
Correlative sparsity: convergence rate?

Term sparsity: (smart) solution extraction

Ideal sparsity: tensor ranks?

Numerical conditioning of sparse SDP relaxations?

Further topics



Correlative sparsity: convergence rate?

Term sparsity: (smart) solution extraction

Ideal sparsity: tensor ranks?

Numerical conditioning of sparse SDP relaxations?

💡 Tons of applications!

Take-away and advertisement

Why should you do polynomial optimization?

Take-away and advertisement

Why should you do polynomial optimization?

💡 powerful & accurate MODELING tool for many applications

Take-away and advertisement

Why should you do polynomial optimization?

💡 powerful & accurate MODELING tool for many applications

💡 EFFICIENCY guaranteed on structured applications: deep learning, quantum information, energy networks

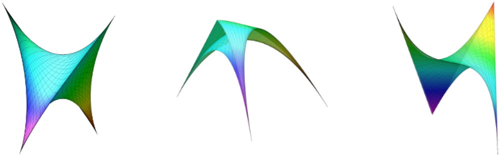
Take-away and advertisement

Why should you do polynomial optimization?

💡 powerful & accurate MODELING tool for many applications

💡 EFFICIENCY guaranteed on structured applications: deep learning, quantum information, energy networks

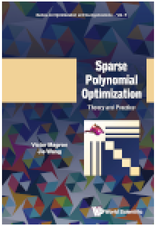
Complementary SYMMETRY exploiting framework on Thursday by Tobias



Thank you for your attention!

`https://homepages.laas.fr/vmagron`

Magron & Wang. Sparse polynomial optimization: theory and practice. To appear in *Series on Optimization and Its Applications, World Scientific Press*, 2022



Korda, Laurent, Magron & Steenkamp. Exploiting ideal-sparsity in the generalized moment problem with application to matrix factorization ranks, arxiv:2209.09573

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

References

- 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
- 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, Computational Optimization and Applications, 2022

References

Chen, Lasserre, Magron & Pauwels. Semialgebraic Optimization for Bounding Lipschitz Constants of ReLU Networks. NeurIPS 2020

Chen, Lasserre, Magron & Pauwels. Semialgebraic Representation of Monotone Deep Equilibrium Models and Applications to Certification. NeuRIPS 2021

Mai, Lasserre & Magron. A sparse version of Reznick's Positivstellensatz, Math OR, 2022

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
  `TSSOS`
- 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. Trans. Math. Soft., 2022
- 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