Nonnegativity, Sparsity and Polynomial Optimization

Jie Wang

Joint work with Victor Magron and Jean B. Lasserre

University of Hong Kong

21 April, 2022

Outline

1. Background on polynomial nonnegativity
   - SOS decomposition
   - SONC decomposition
2. Sparsity in the moment-SOS hierarchy
   - Correlative sparsity (CS)
   - Term sparsity (TS)
3. SONC decomposition and second order cone representation
4. Numerical experiments

Outline

1. Background on polynomial nonnegativity
   - SOS decomposition
   - SONC decomposition
2. Sparsity in the moment-SOS hierarchy
   - Correlative sparsity (CS)
   - Term sparsity (TS)
3. SONC decomposition and second order cone representation
4. Numerical experiments

Outline

1. Background on polynomial nonnegativity
   - SOS decomposition
   - SONC decomposition
2. Sparsity in the moment-SOS hierarchy
   - Correlative sparsity (CS)
   - Term sparsity (TS)
3. SONC decomposition and second order cone representation
4. Numerical experiments

Outline

1. Background on polynomial nonnegativity
   - SOS decomposition
   - SONC decomposition
2. Sparsity in the moment-SOS hierarchy
   - Correlative sparsity (CS)
   - Term sparsity (TS)
3. SONC decomposition and second order cone representation
4. Numerical experiments

Certify polynomial nonnegativity

Problem

Given a multivariate polynomial f, decide if it is nonnegative and if so, provide a certificate of nonnegativity.

- A central problem in real algebraic geometry
- Widely appear in numerous fields
- Closely related to polynomial optimization
- NP-hard in general

Certify polynomial nonnegativity

Problem

Given a multivariate polynomial f, decide if it is nonnegative and if so, provide a certificate of nonnegativity.

- A central problem in real algebraic geometry
- Widely appear in numerous fields
- Closely related to polynomial optimization
- NP-hard in general

Certify polynomial nonnegativity

Problem

Given a multivariate polynomial f, decide if it is nonnegative and if so, provide a certificate of nonnegativity.

- A central problem in real algebraic geometry
- Widely appear in numerous fields
- Closely related to polynomial optimization
- NP-hard in general

Certify polynomial nonnegativity

Problem

Given a multivariate polynomial f, decide if it is nonnegative and if so, provide a certificate of nonnegativity.

- A central problem in real algebraic geometry
- Widely appear in numerous fields
- Closely related to polynomial optimization
- NP-hard in general

Certify polynomial nonnegativity

Problem

Given a multivariate polynomial f, decide if it is nonnegative and if so, provide a certificate of nonnegativity.

- A central problem in real algebraic geometry
- Widely appear in numerous fields
- Closely related to polynomial optimization
- NP-hard in general

SOS decomposition

- SOS (sum of squares) decomposition:

$$f = f_1^2 + \cdots + f_t^2 \quad \rightsquigarrow \quad f \text{ is nonnegative}$$

Example: $f = 1 + 2x + 2x^2 + 2xy + y^2 = (1 + x)^2 + (x + y)^2$

- Hilbert, 1888:

"nonnegative polynomials = SOS" $\Leftrightarrow n = 1;\ d = 2;\ n = 2, d = 4$

- Artin, 1927: "nonnegative polynomials = rational SOS"
- Blekherman, 2006: "nonnegative polynomials $\gg$ SOS", $n \to \infty$
- Motzkin's polynomial: $M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$

SOS decomposition

- SOS (sum of squares) decomposition:

$$f = f_1^2 + \cdots + f_t^2 \quad \rightsquigarrow \quad f \text{ is nonnegative}$$

Example: $f = 1 + 2x + 2x^2 + 2xy + y^2 = (1 + x)^2 + (x + y)^2$

- Hilbert, 1888:

"nonnegative polynomials = SOS" $\Leftrightarrow n = 1;\ d = 2;\ n = 2, d = 4$

- Artin, 1927: "nonnegative polynomials = rational SOS"
- Blekherman, 2006: "nonnegative polynomials $\gg$ SOS", $n \to \infty$
- Motzkin's polynomial: $M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$

SOS decomposition

- SOS (sum of squares) decomposition:

$$f = f_1^2 + \cdots + f_t^2 \quad \rightsquigarrow \quad f \text{ is nonnegative}$$

Example: $f = 1 + 2x + 2x^2 + 2xy + y^2 = (1 + x)^2 + (x + y)^2$

- Hilbert, 1888:

"nonnegative polynomials = SOS" $\Leftrightarrow n = 1;\ d = 2;\ n = 2, d = 4$

- Artin, 1927: "nonnegative polynomials = rational SOS"
- Blekherman, 2006: "nonnegative polynomials $\gg$ SOS", $n \to \infty$
- Motzkin's polynomial: $M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$

SOS decomposition

- SOS (sum of squares) decomposition:

$$f = f_1^2 + \cdots + f_t^2 \quad \rightsquigarrow \quad f \text{ is nonnegative}$$

Example: $f = 1 + 2x + 2x^2 + 2xy + y^2 = (1 + x)^2 + (x + y)^2$

- Hilbert, 1888:

"nonnegative polynomials = SOS" $\Leftrightarrow n = 1;\ d = 2;\ n = 2, d = 4$

- Artin, 1927: "nonnegative polynomials = rational SOS"
- Blekherman, 2006: "nonnegative polynomials $\gg$ SOS", $n \to \infty$
- Motzkin's polynomial: $M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$

SOS decomposition

- SOS (sum of squares) decomposition:

$$f = f_1^2 + \cdots + f_t^2 \quad \rightsquigarrow \quad f \text{ is nonnegative}$$

Example: $f = 1 + 2x + 2x^2 + 2xy + y^2 = (1 + x)^2 + (x + y)^2$

- Hilbert, 1888:

"nonnegative polynomials = SOS" $\Leftrightarrow n = 1$; $d = 2$; $n = 2, d = 4$

- Artin, 1927: "nonnegative polynomials = rational SOS"
- Blekherman, 2006: "nonnegative polynomials $\gg$ SOS", $n \to \infty$
- Motzkin's polynomial: $M(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$

Gram matrix and semidefinite programming

- f: $2d$, $v_d = [1, x_1, \ldots, x_n, x_1^d, \ldots, x_n^d]$
- f admits an SOS decomposition $\iff$ there exists a PSD matrix G s.t.

$$f = v_d \cdot G \cdot v_d^{\intercal} \quad \rightsquigarrow \quad \text{SDP}$$

- G is called a Gram matrix of f, which is of size $\binom{n+d}{n}$

Gram matrix and semidefinite programming

- f: $2d$, $v_d = [1, x_1, \ldots, x_n, x_1^d, \ldots, x_n^d]$
- f admits an SOS decomposition $\iff$ there exists a PSD matrix G s.t.

$$f = v_d \cdot G \cdot v_d^\intercal \quad \rightsquigarrow \quad \text{SDP}$$

- G is called a Gram matrix of f, which is of size $\binom{n+d}{n}$

Gram matrix and semidefinite programming

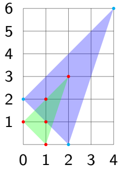
- f: $2d$, $v_d = [1, x_1, \ldots, x_n, x_1^d, \ldots, x_n^d]$
- f admits an SOS decomposition $\iff$ there exists a PSD matrix G s.t.

$$f = v_d \cdot G \cdot v_d^{\intercal} \quad \rightsquigarrow \quad \text{SDP}$$

- G is called a Gram matrix of f, which is of size $\binom{n+d}{n}$

Structural SOS decomposition

- Newton polytope: $f = \sum f_i^2 \Longrightarrow \mathrm{New}(f_i) \subseteq \frac{1}{2}\mathrm{New}(f)$



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

- correlative sparsity

$$f(\mathbf{x}) = f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) \in \mathrm{SOS} \rightsquigarrow f_1(\mathbf{x}_1) \in \mathrm{SOS}, f_2(\mathbf{x}_2) \in \mathrm{SOS}$$

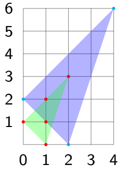
- term sparsity

$$\mathbf{x}^{\beta} \cdot \mathbf{x}^{\gamma} \notin \mathrm{supp}(f), \beta + \gamma \notin (2\mathbb{N})^n \quad \rightsquigarrow \quad G_{\beta\gamma} = 0$$

- symmetry

Structural SOS decomposition

- Newton polytope: $f = \sum f_i^2 \Longrightarrow \mathrm{New}(f_i) \subseteq \frac{1}{2}\mathrm{New}(f)$



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

- correlative sparsity

$$f(\mathbf{x}) = f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) \in \mathrm{SOS} \rightsquigarrow f_1(\mathbf{x}_1) \in \mathrm{SOS}, f_2(\mathbf{x}_2) \in \mathrm{SOS}$$

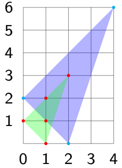
- term sparsity

$$\mathbf{x}^{\beta} \cdot \mathbf{x}^{\gamma} \notin \mathrm{supp}(f), \beta + \gamma \notin (2\mathbb{N})^n \quad \rightsquigarrow \quad G_{\beta\gamma} = 0$$

- symmetry

Structural SOS decomposition

- Newton polytope: $f = \sum f_i^2 \Longrightarrow \mathrm{New}(f_i) \subseteq \frac{1}{2}\mathrm{New}(f)$



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

- correlative sparsity

$$f(\mathbf{x}) = f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) \in \mathrm{SOS} \rightsquigarrow f_1(\mathbf{x}_1) \in \mathrm{SOS}, f_2(\mathbf{x}_2) \in \mathrm{SOS}$$

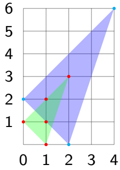
- term sparsity

$$\mathbf{x}^{\boldsymbol{\beta}} \cdot \mathbf{x}^{\gamma} \notin \mathrm{supp}(f), \boldsymbol{\beta} + \boldsymbol{\gamma} \notin (2\mathbb{N})^n \quad \rightsquigarrow \quad G_{\boldsymbol{\beta}\boldsymbol{\gamma}} = 0$$

- symmetry

Structural SOS decomposition

- Newton polytope: $f = \sum f_i^2 \Longrightarrow \mathrm{New}(f_i) \subseteq \frac{1}{2}\mathrm{New}(f)$



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

- correlative sparsity

$$f(\mathbf{x}) = f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) \in \mathrm{SOS} \rightsquigarrow f_1(\mathbf{x}_1) \in \mathrm{SOS}, f_2(\mathbf{x}_2) \in \mathrm{SOS}$$

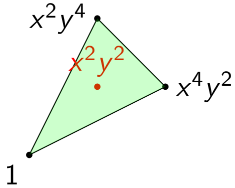
- term sparsity

$$\mathbf{x}^{\boldsymbol{\beta}} \cdot \mathbf{x}^{\gamma} \notin \mathrm{supp}(f), \boldsymbol{\beta} + \boldsymbol{\gamma} \notin (2\mathbb{N})^n \quad \rightsquigarrow \quad G_{\boldsymbol{\beta}\boldsymbol{\gamma}} = 0$$

- symmetry

SONC (sum of nonnegative circuits) decomposition

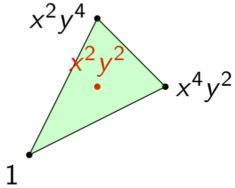
- $M(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ (arithmetic-geometric mean inequality $\Rightarrow$ nonnegativity)



- circuit polynomial: $f = \sum_{\alpha \in \mathscr{A}} c_\alpha \mathbf{x}^\alpha - d_\beta \mathbf{x}^\beta$, $\alpha \in (2\mathbb{N})^n$, $c_\alpha > 0$, $\mathscr{A}$ the vertex set of a simplex, $\beta \in \mathrm{conv}(\mathscr{A})^\circ$
- SONC decomposition: $f = f_1 + \cdots + f_t$ with each f_i being a nonnegative circuit polynomial

SONC (sum of nonnegative circuits) decomposition

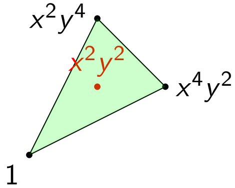
- $M(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ (arithmetic-geometric mean inequality $\Rightarrow$ nonnegativity)



- circuit polynomial: $f = \sum_{\boldsymbol{\alpha}\in\mathscr{A}} c_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}} - d_{\boldsymbol{\beta}}\mathbf{x}^{\boldsymbol{\beta}}$, $\boldsymbol{\alpha} \in (2\mathbb{N})^n$, $c_{\boldsymbol{\alpha}} > 0$, $\mathscr{A}$ the vertex set of a simplex, $\boldsymbol{\beta} \in \mathrm{conv}(\mathscr{A})^\circ$
- SONC decomposition: $f = f_1 + \cdots + f_t$ with each f_i being a nonnegative circuit polynomial

SONC (sum of nonnegative circuits) decomposition

- $M(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$ (arithmetic-geometric mean inequality $\Rightarrow$ nonnegativity)



- circuit polynomial: $f = \sum_{\boldsymbol{\alpha}\in\mathscr{A}} c_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}} - d_{\boldsymbol{\beta}}\mathbf{x}^{\boldsymbol{\beta}}$, $\boldsymbol{\alpha} \in (2\mathbb{N})^n$, $c_{\boldsymbol{\alpha}} > 0$, $\mathscr{A}$ the vertex set of a simplex, $\boldsymbol{\beta} \in \mathrm{conv}(\mathscr{A})^\circ$
- SONC decomposition: $f = f_1 + \cdots + f_t$ with each f_i being a nonnegative circuit polynomial

Sufficient conditions for SONC decompositions

Theorem (Wang, 2022)

Suppose that f is a nonnegative polynomial with exactly one negative term. Then f admits a SONC decomposition.

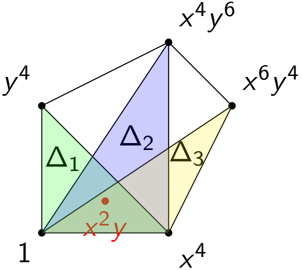
- $f = 1 + x^4 + y^4 + x^6y^4 + x^4y^6 - x^2y$

Sufficient conditions for SONC decompositions

Theorem (Wang, 2022)

Suppose that f is a nonnegative polynomial with exactly one negative term. Then f admits a SONC decomposition.

- $f = 1 + x^4 + y^4 + x^6y^4 + x^4y^6 - x^2y$



SONC decompositions preserve term sparsity

Theorem (Wang, 2022)

Suppose f is a SONC polynomial. Then f admits a SONC decomposition:

$$f = \sum_{\mathrm{supp}(f_i)\subseteq\mathrm{supp}(f)} f_i,$$

where each f_i is a nonnegative circuit polynomial. Moreover, we can further assume that there is no cancellation occurring in the above decomposition.

Sparsity in the moment-SOS hierarchy

Polynomial optimization

- Polynomial optimization problem:

$$f^* := \begin{cases} \inf & f \\ \text{s.t.} & g_j \geq 0, \quad j = 1, \ldots, m \\ & (h_i = 0, \quad i = 1, \ldots, m') \end{cases}$$

- non-convex, NP-hard
- Optimal power flow, computer vision, neural networks, signal processing, quantum information......

Polynomial optimization

- Polynomial optimization problem:

$$f^* := \begin{cases} \inf & f \\ \text{s.t.} & g_j \geq 0, \quad j = 1, \ldots, m \\ & (h_i = 0, \quad i = 1, \ldots, m') \end{cases}$$

- non-convex, NP-hard
- Optimal power flow, computer vision, neural networks, signal processing, quantum information......

Polynomial optimization

- Polynomial optimization problem:

$$f^* := \begin{cases} \inf & f \\ \text{s.t.} & g_j \geq 0, \quad j = 1, \ldots, m \\ & (h_i = 0, \quad i = 1, \ldots, m') \end{cases}$$

- non-convex, NP-hard
- Optimal power flow, computer vision, neural networks, signal processing, quantum information......

The goals

- Compute the global optimal value
- Certify global optimality
- Extract global optimal solutions
- Approximate the global optimal value if the exact computation is expensive/impossible/unnecessary

The moment-SOS hierarchy (also known as Lasserre's hierarchy) is a well-established tool to handle POPs and achieve the above goals.

The goals

- Compute the global optimal value
- Certify global optimality
- Extract global optimal solutions
- Approximate the global optimal value if the exact computation is expensive/impossible/unnecessary

The moment-SOS hierarchy (also known as Lasserre's hierarchy) is a well-established tool to handle POPs and achieve the above goals.

The goals

- Compute the global optimal value
- Certify global optimality
- Extract global optimal solutions
- Approximate the global optimal value if the exact computation is expensive/impossible/unnecessary

The moment-SOS hierarchy (also known as Lasserre's hierarchy) is a well-established tool to handle POPs and achieve the above goals.

The goals

- Compute the global optimal value
- Certify global optimality
- Extract global optimal solutions
- Approximate the global optimal value if the exact computation is expensive/impossible/unnecessary

The moment-SOS hierarchy (also known as Lasserre's hierarchy) is a well-established tool to handle POPs and achieve the above goals.

The goals

- Compute the global optimal value
- Certify global optimality
- Extract global optimal solutions
- Approximate the global optimal value if the exact computation is expensive/impossible/unnecessary

The moment-SOS hierarchy (also known as Lasserre's hierarchy) is a well-established tool to handle POPs and achieve the above goals.

Moment matrix and localizing matrix

- $\mathbb{N}_r^n := \{\boldsymbol{\beta} = (\beta_i) \in \mathbb{N}^n \mid \sum_{i=1}^n \beta_i \le r\}$
- r-th order moment matrix $M_r(\mathbf{y})$:

$$[M_r(\mathbf{y})]_{\boldsymbol{\beta}\boldsymbol{\gamma}} := y_{\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \forall \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{N}_r^n$$

- Given $g = \sum_{\boldsymbol{\alpha}} g_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$, r-th order localizing matrix $M_r(g\mathbf{y})$:

$$[M_r(g\mathbf{y})]_{\boldsymbol{\beta}\boldsymbol{\gamma}} := \sum_{\boldsymbol{\alpha}} g_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \forall \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{N}_r^n$$

- $\mathbf{x} = x$, $g = 1 - x^2$:

$$M_2(\mathbf{y}) = \begin{matrix} & 1 & x & x^2 \\ 1 & y_0 & y_1 & y_2 \\ x & y_1 & y_2 & y_3 \\ x^2 & y_2 & y_3 & y_4 \end{matrix}, \quad M_1(g\mathbf{y}) = \begin{matrix} & 1 & x \\ 1 & y_0 - y_2 & y_1 - y_3 \\ x & y_1 - y_3 & y_2 - y_4 \end{matrix}$$

Moment matrix and localizing matrix

- $\mathbb{N}^n_r := \{\boldsymbol{\beta} = (\beta_i) \in \mathbb{N}^n \mid \sum_{i=1}^n \beta_i \le r\}$
- r-th order moment matrix $M_r(\mathbf{y})$:

$$[M_r(\mathbf{y})]_{\boldsymbol{\beta}\boldsymbol{\gamma}} := y_{\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \forall \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{N}^n_r$$

- Given $g = \sum_{\alpha} g_{\alpha} \mathbf{x}^{\alpha}$, r-th order localizing matrix $M_r(g\mathbf{y})$:

$$[M_r(g\mathbf{y})]_{\boldsymbol{\beta}\boldsymbol{\gamma}} := \sum_{\alpha} g_{\alpha} y_{\alpha+\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \forall \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{N}^n_r$$

- $\mathbf{x} = x$, $g = 1 - x^2$:

$$M_2(\mathbf{y}) = \begin{array}{c} \\ 1 \\ x \\ x^2 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & x & x^2 \end{array} \\ \begin{pmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{pmatrix} \end{array}, \quad M_1(g\mathbf{y}) = \begin{array}{c} \\ 1 \\ x \end{array} \begin{array}{c} \begin{array}{cc} 1 & x \end{array} \\ \begin{pmatrix} y_0 - y_2 & y_1 - y_3 \\ y_1 - y_3 & y_2 - y_4 \end{pmatrix} \end{array}$$

Moment matrix and localizing matrix

- $\mathbb{N}^n_r := \{\boldsymbol{\beta} = (\beta_i) \in \mathbb{N}^n \mid \sum_{i=1}^n \beta_i \le r\}$
- r-th order moment matrix $M_r(\mathbf{y})$:

$$[M_r(\mathbf{y})]_{\boldsymbol{\beta}\boldsymbol{\gamma}} := y_{\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \forall \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{N}^n_r$$

- Given $g = \sum_{\boldsymbol{\alpha}} g_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$, r-th order localizing matrix $M_r(g\mathbf{y})$:

$$[M_r(g\mathbf{y})]_{\boldsymbol{\beta}\boldsymbol{\gamma}} := \sum_{\boldsymbol{\alpha}} g_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \forall \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{N}^n_r$$

- $\mathbf{x} = x$, $g = 1 - x^2$:

$$M_2(\mathbf{y}) = \begin{matrix} & 1 & x & x^2 \\ 1 & y_0 & y_1 & y_2 \\ x & y_1 & y_2 & y_3 \\ x^2 & y_2 & y_3 & y_4 \end{matrix}, \quad M_1(g\mathbf{y}) = \begin{matrix} & 1 & x \\ 1 & y_0 - y_2 & y_1 - y_3 \\ x & y_1 - y_3 & y_2 - y_4 \end{matrix}$$

Moment matrix and localizing matrix

- $\mathbb{N}^n_r := \{\boldsymbol{\beta} = (\beta_i) \in \mathbb{N}^n \mid \sum_{i=1}^n \beta_i \le r\}$
- r-th order moment matrix $M_r(\mathbf{y})$:

$$[M_r(\mathbf{y})]_{\boldsymbol{\beta}\boldsymbol{\gamma}} := y_{\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \forall \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{N}^n_r$$

- Given $g = \sum_{\boldsymbol{\alpha}} g_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$, r-th order localizing matrix $M_r(g\mathbf{y})$:

$$[M_r(g\mathbf{y})]_{\boldsymbol{\beta}\boldsymbol{\gamma}} := \sum_{\boldsymbol{\alpha}} g_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}}, \quad \forall \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{N}^n_r$$

- $\mathbf{x} = x$, $g = 1 - x^2$:

$$M_2(\mathbf{y}) = \begin{array}{c} \\ 1 \\ x \\ x^2 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & x & x^2 \end{array} \\ \begin{pmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{pmatrix} \end{array}, \quad M_1(g\mathbf{y}) = \begin{array}{c} \\ 1 \\ x \end{array} \begin{array}{c} \begin{array}{cc} 1 & x \end{array} \\ \begin{pmatrix} y_0 - y_2 & y_1 - y_3 \\ y_1 - y_3 & y_2 - y_4 \end{pmatrix} \end{array}$$

Moment relaxation

- Moment relaxation:

$$\theta_r := \begin{cases} \inf & L_{\mathbf{y}}(f) \\ \text{s.t.} & M_r(\mathbf{y}) \succeq 0, \\ & M_{r-d_j}(g_j\mathbf{y}) \succeq 0, \quad j = 1, \ldots, m, \\ & y_{\mathbf{0}} = 1. \end{cases}$$

SOS relaxation

- $S = \{\mathbf{x} \in \mathbb{R}^n \mid g_1 \geq 0, \ldots, g_m \geq 0\}$
- Dual to the polynomial optimization problem:

$$f^* = \sup_{\lambda}\{\lambda : f(\mathbf{x}) - \lambda \geq 0, \forall \mathbf{x} \in S\}$$

- $P_S(\mathbf{x}) := \{g(\mathbf{x}) \in \mathbb{R}[\mathbf{x}] \mid g(\mathbf{x}) \geq 0 \text{ over } S\} \rightsquigarrow$ intractable
- Approximate $P_S(\mathbf{x})$ by more tractable subsets $\rightsquigarrow$ SOS, SONC

SOS relaxation

- $S = \{\mathbf{x} \in \mathbb{R}^n \mid g_1 \geq 0, \ldots, g_m \geq 0\}$
- Dual to the polynomial optimization problem:

$$f^* = \sup_{\lambda} \{\lambda : f(\mathbf{x}) - \lambda \geq 0, \forall \mathbf{x} \in S\}$$

- $P_S(\mathbf{x}) := \{g(\mathbf{x}) \in \mathbb{R}[\mathbf{x}] \mid g(\mathbf{x}) \geq 0 \text{ over } S\} \rightsquigarrow$ intractable
- Approximate $P_S(\mathbf{x})$ by more tractable subsets $\rightsquigarrow$ SOS, SONC

SOS relaxation

- $S = \{\mathbf{x} \in \mathbb{R}^n \mid g_1 \geq 0, \ldots, g_m \geq 0\}$
- Dual to the polynomial optimization problem:

$$f^* = \sup_{\lambda} \{\lambda : f(\mathbf{x}) - \lambda \geq 0, \forall \mathbf{x} \in S\}$$

- $P_S(\mathbf{x}) := \{g(\mathbf{x}) \in \mathbb{R}[\mathbf{x}] \mid g(\mathbf{x}) \geq 0 \text{ over } S\}$ $\rightsquigarrow$ intractable
- Approximate $P_S(\mathbf{x})$ by more tractable subsets $\rightsquigarrow$ SOS, SONC

Quadratic module

- $\Sigma(\mathbf{x}) := \{f \in \mathbb{R}[\mathbf{x}] \mid f = \sum_i f_i^2, f_i \in \mathbb{R}[\mathbf{x}]\}$
- Quadratic module: Given $\mathbf{g} = \{g_j\}_{j=1}^m \subseteq \mathbb{R}[\mathbf{x}]$,

$$\mathcal{Q}(\mathbf{g}) := \left\{\sigma_0 + \sum_{j=1}^{m} \sigma_j g_j \mid \sigma_j \in \Sigma(\mathbf{x}), j = 0, 1, \ldots, m\right\} \subseteq P_S(\mathbf{x})$$

- Truncated quadratic module:

$$\mathcal{Q}(\mathbf{g})_{2r} := \left\{\sigma_0 + \sum_{j=1}^{m} \sigma_j g_j \mid \sigma_j \in \Sigma(\mathbf{x}), \deg(\sigma_j g_j) \le 2r, j = 0, 1, \ldots, m\right\}$$

Quadratic module

- $\Sigma(\mathbf{x}) := \{f \in \mathbb{R}[\mathbf{x}] \mid f = \sum_i f_i^2, f_i \in \mathbb{R}[\mathbf{x}]\}$
- Quadratic module: Given $\mathbf{g} = \{g_j\}_{j=1}^m \subseteq \mathbb{R}[\mathbf{x}]$,

$$\mathcal{Q}(\mathbf{g}) := \left\{\sigma_0 + \sum_{j=1}^m \sigma_j g_j \mid \sigma_j \in \Sigma(\mathbf{x}), j = 0, 1, \ldots, m\right\} \subseteq P_S(\mathbf{x})$$

- Truncated quadratic module:

$$\mathcal{Q}(\mathbf{g})_{2r} := \left\{\sigma_0 + \sum_{j=1}^m \sigma_j g_j \mid \sigma_j \in \Sigma(\mathbf{x}), \deg(\sigma_j g_j) \leq 2r, j = 0, 1, \ldots, m\right\}$$

Quadratic module

- $\Sigma(\mathbf{x}) := \{f \in \mathbb{R}[\mathbf{x}] \mid f = \sum_i f_i^2, f_i \in \mathbb{R}[\mathbf{x}]\}$
- Quadratic module: Given $\mathbf{g} = \{g_j\}_{j=1}^m \subseteq \mathbb{R}[\mathbf{x}]$,

$$\mathcal{Q}(\mathbf{g}) := \left\{\sigma_0 + \sum_{j=1}^m \sigma_j g_j \mid \sigma_j \in \Sigma(\mathbf{x}), j = 0, 1, \ldots, m\right\} \subseteq P_S(\mathbf{x})$$

- Truncated quadratic module:

$$\mathcal{Q}(\mathbf{g})_{2r} := \left\{\sigma_0 + \sum_{j=1}^m \sigma_j g_j \mid \sigma_j \in \Sigma(\mathbf{x}), \deg(\sigma_j g_j) \leq 2r, j = 0, 1, \ldots, m\right\}$$

Dual SOS relaxation

- Dual SOS relaxation:

$$\theta_r^* := \begin{cases} \sup & \lambda \\ \text{s.t.} & f - \lambda \in \mathcal{Q}(\mathbf{g})_{2r}. \end{cases}$$

$$\Updownarrow$$

$$\theta_r^* := \begin{cases} \sup & \lambda \\ \text{s.t.} & f - \lambda = \sigma_0 + \sum_{j=1}^m \sigma_j g_j, \\ & \sigma_0, \sigma_1, \ldots, \sigma_m \in \Sigma(\mathbf{x}), \\ & \deg(\sigma_0) \le 2r, \deg(\sigma_j g_j) \le 2r, j = 1, \ldots, m. \end{cases}$$

The moment-SOS hierarchy

$$
\begin{array}{ccccc}
 & & f^* & & \\
 & \rotatebox{-45}{$\leq$} & & \rotatebox{45}{$\geq$} & \\
 & \vdots & & \vdots & \\
 & \leq & & \leq & \\
\text{(moment relaxation)} & \theta_r & \text{“}=\text{”} & \theta_r^* & \text{(SOS relaxation)} \\
 & \leq & & \leq & \\
 & \vdots & & \vdots & \\
 & \leq & & \leq & \\
 & \theta_{\underline{r}} & \text{“}=\text{”} & \theta_{\underline{r}}^* &
\end{array}
$$

Asymptotical convergence and finite convergence

- Archimedean's condition: there exists $N > 0$ s.t. $N - ||\mathbf{x}||^2 \in \mathcal{Q}(\mathbf{g})$

➤ $\theta_r \uparrow f^*$ and $\theta_r^* \uparrow f^*$ when $r \to \infty$ (Lasserre, 2001);

➤ Finite convergence happens generically (Nie, 2014);

➤ We can verify global optimality by the so-called rank condition (flat extension/truncation);

➤ We can easily extract minimizers when the rank condition is satisfied.

Asymptotical convergence and finite convergence

- Archimedean's condition: there exists $N > 0$ s.t. $N - ||\mathbf{x}||^2 \in \mathcal{Q}(\mathbf{g})$

  ➤ $\theta_r \uparrow f^*$ and $\theta_r^* \uparrow f^*$ when $r \to \infty$ (Lasserre, 2001);

  ➤ Finite convergence happens generically (Nie, 2014);

  ➤ We can verify global optimality by the so-called rank condition (flat extension/truncation);

  ➤ We can easily extract minimizers when the rank condition is satisfied.

Asymptotical convergence and finite convergence

- Archimedean's condition: there exists $N > 0$ s.t. $N - ||\mathbf{x}||^2 \in \mathcal{Q}(\mathbf{g})$
  - ➤ $\theta_r \uparrow f^*$ and $\theta_r^* \uparrow f^*$ when $r \to \infty$ (Lasserre, 2001);
  - ➤ Finite convergence happens generically (Nie, 2014);
  - ➤ We can verify global optimality by the so-called rank condition (flat extension/truncation);
  - ➤ We can easily extract minimizers when the rank condition is satisfied.

Asymptotical convergence and finite convergence

- Archimedean's condition: there exists $N > 0$ s.t. $N - ||\mathbf{x}||^2 \in \mathcal{Q}(\mathbf{g})$

➤ $\theta_r \uparrow f^*$ and $\theta_r^* \uparrow f^*$ when $r \to \infty$ (Lasserre, 2001);

➤ Finite convergence happens generically (Nie, 2014);

➤ We can verify global optimality by the so-called rank condition (flat extension/truncation);

➤ We can easily extract minimizers when the rank condition is satisfied.

Asymptotical convergence and finite convergence

- Archimedean's condition: there exists $N > 0$ s.t. $N - ||\mathbf{x}||^2 \in \mathcal{Q}(\mathbf{g})$

  ➤ $\theta_r \uparrow f^*$ and $\theta_r^* \uparrow f^*$ when $r \to \infty$ (Lasserre, 2001);

  ➤ Finite convergence happens generically (Nie, 2014);

  ➤ We can verify global optimality by the so-called rank condition (flat extension/truncation);

  ➤ We can easily extract minimizers when the rank condition is satisfied.

Scalability issue

- The size of SDP (considering the SOS problem) at relaxation order r:
  1. maximal size of PSD matrices: $\binom{n+r}{r}$
  2. number of equality constraints: $\binom{n+2r}{2r}$
- $r = 2$, $n \leq 30$ (Mosek)
- Exploiting structure:
  - quotient ring
  - symmetry
  - sparsity

Scalability issue

- The size of SDP (considering the SOS problem) at relaxation order r:
  1. maximal size of PSD matrices: $\binom{n+r}{r}$
  2. number of equality constraints: $\binom{n+2r}{2r}$
- $r = 2$, $n \leq 30$ (Mosek)
- Exploiting structure:
  - quotient ring
  - symmetry
  - sparsity

Scalability issue

- The size of SDP (considering the SOS problem) at relaxation order r:
  1. maximal size of PSD matrices: $\binom{n+r}{r}$
  2. number of equality constraints: $\binom{n+2r}{2r}$
- $r = 2$, $n \leq 30$ (Mosek)
- Exploiting structure:
  - quotient ring
  - symmetry
  - sparsity

Correlative sparsity (Waki et al., 2006)

- Correlative sparsity pattern (csp) graph $G^{\text{csp}}(V, E)$:
  - $V := \{x_1, \ldots, x_n\}$
  - $\{x_i, x_j\} \in E \Longleftrightarrow x_i, x_j$ appear in the same term of f or appear in the same constraint polynomial g_k
- For each maximal clique of the csp graph $G^{\text{csp}}(V, E)$:

$$I_k \longmapsto M_r(\mathbf{y}, I_k), M_{r-d_j}(g_j\mathbf{y}, I_k)$$

Correlative sparsity (Waki et al., 2006)

- Correlative sparsity pattern (csp) graph $G^{\mathrm{csp}}(V, E)$:

  ➤ $V := \{x_1, \ldots, x_n\}$

  ➤ $\{x_i, x_j\} \in E \Longleftrightarrow x_i, x_j$ appear in the same term of f or appear in the same constraint polynomial g_k

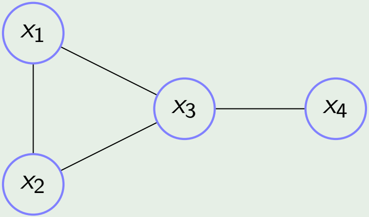
- For each maximal clique of the csp graph $G^{\mathrm{csp}}(V, E)$:

$$I_k \longmapsto M_r(\mathbf{y}, I_k), M_{r-d_j}(g_j\mathbf{y}, I_k)$$

Correlative sparsity

Example

$f = x_1^4 + x_1x_2^2 + x_2x_3 + x_3^2x_4^2$, $g_1 = 1 - x_1^2 - x_2^2 - x_3^2$, $g_2 = 1 - x_3x_4$



There are two maximal cliques: $\{x_1, x_2, x_3\}$ and $\{x_3, x_4\}$

The moment-SOS hierarchy based on correlative sparsity

- If the csp graph is chordal (otherwise we need a chordal extension), then the moment-SOS hierarchy based on correlative sparsity shares the same convergence as the standard one;
- We can still verify global optimality by the (adapted) rank condition;
- We can still extract global minimizers if certain rank conditions are satisfied;
- Significantly improve scalability if the sizes of maximal cliques of the csp graph are small (e.g. ≤ 10).

The moment-SOS hierarchy based on correlative sparsity

- If the csp graph is chordal (otherwise we need a chordal extension), then the moment-SOS hierarchy based on correlative sparsity shares the same convergence as the standard one;
- We can still verify global optimality by the (adapted) rank condition;
- We can still extract global minimizers if certain rank conditions are satisfied;
- Significantly improve scalability if the sizes of maximal cliques of the csp graph are small (e.g. ≤ 10).

The moment-SOS hierarchy based on correlative sparsity

- If the csp graph is chordal (otherwise we need a chordal extension), then the moment-SOS hierarchy based on correlative sparsity shares the same convergence as the standard one;
- We can still verify global optimality by the (adapted) rank condition;
- We can still extract global minimizers if certain rank conditions are satisfied;
- Significantly improve scalability if the sizes of maximal cliques of the csp graph are small (e.g. ≤ 10).

The moment-SOS hierarchy based on correlative sparsity

- If the csp graph is chordal (otherwise we need a chordal extension), then the moment-SOS hierarchy based on correlative sparsity shares the same convergence as the standard one;
- We can still verify global optimality by the (adapted) rank condition;
- We can still extract global minimizers if certain rank conditions are satisfied;
- Significantly improve scalability if the sizes of maximal cliques of the csp graph are small (e.g. ≤ 10).

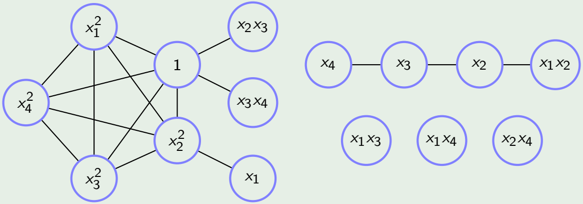
Term sparsity (Wang, Magron, and Lasserre, 2021)

- Term sparsity pattern (tsp) graph $G^{\text{tsp}}(V, E)$:
  - $V := v_r = \{1, x_1, \ldots, x_n, x_1^r, \ldots, x_n^r\}$
  - $\{\mathbf{x}^{\boldsymbol{\alpha}}, \mathbf{x}^{\boldsymbol{\beta}}\} \in E \iff \mathbf{x}^{\boldsymbol{\alpha}} \cdot \mathbf{x}^{\boldsymbol{\beta}} = \mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \in \mathrm{supp}(f) \cup \bigcup_{j=1}^{m} \mathrm{supp}(g_j) \cup v_r^2$

Term sparsity

Example

$f = x_1^4 + x_1x_2^2 + x_2x_3 + x_3^2x_4^2$, $g_1 = 1 - x_1^2 - x_2^2 - x_3^2$, $g_2 = 1 - x_3x_4$



Term sparsity

- Assume $(G^{\rm tsp})'$ is a chordal extension of $G^{\rm tsp}$ with maximal cliques: $C_1, \ldots, C_t$

$$C_i \longmapsto M_{C_i}(\mathbf{y}), \quad i = 1, \ldots, t$$

- Decompose the moment matrix:

$$M_r(\mathbf{y}) \succeq 0 \longrightarrow M_{C_i}(\mathbf{y}) \succeq 0, \quad i = 1, \ldots, t$$

- Decompose the localizing matrix $M_{r-d_j}(\mathbf{y})$, $j = 1, \ldots, m$

Term sparsity

- Assume $(G^{\mathrm{tsp}})'$ is a chordal extension of G^{tsp} with maximal cliques: $C_1, \ldots, C_t$

$$C_i \longmapsto M_{C_i}(\mathbf{y}), \quad i = 1, \ldots, t$$

- Decompose the moment matrix:

$$M_r(\mathbf{y}) \succeq 0 \longrightarrow M_{C_i}(\mathbf{y}) \succeq 0, \quad i = 1, \ldots, t$$

- Decompose the localizing matrix $M_{r-d_j}(\mathbf{y})$, $j = 1, \ldots, m$

Term sparsity

- Assume $(G^{\mathrm{tsp}})'$ is a chordal extension of G^{tsp} with maximal cliques: $C_1, \ldots, C_t$

$$C_i \longmapsto M_{C_i}(\mathbf{y}), \quad i = 1, \ldots, t$$

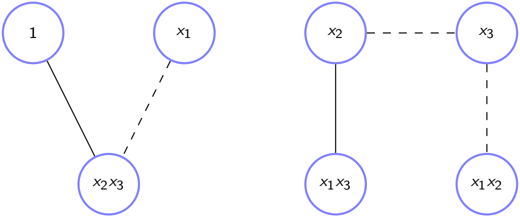
- Decompose the moment matrix:

$$M_r(\mathbf{y}) \succeq 0 \longrightarrow M_{C_i}(\mathbf{y}) \succeq 0, \quad i = 1, \ldots, t$$

- Decompose the localizing matrix $M_{r-d_j}(\mathbf{y})$, $j = 1, \ldots, m$

Extending to an iterative procedure

- support extension: $\mathbf{x}^{\boldsymbol{\beta}'}\mathbf{x}^{\boldsymbol{\gamma}'} = \mathbf{x}^{\boldsymbol{\beta}}\mathbf{x}^{\boldsymbol{\gamma}}$, $\{\mathbf{x}^{\boldsymbol{\beta}}, \mathbf{x}^{\boldsymbol{\gamma}}\} \in E \Rightarrow \{\mathbf{x}^{\boldsymbol{\beta}'}, \mathbf{x}^{\boldsymbol{\gamma}'}\} \in E$

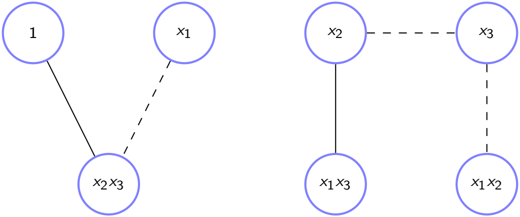


- Iteratively perform support extension and chordal extension:

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

Extending to an iterative procedure

- support extension: $\mathbf{x}^{\boldsymbol{\beta}'}\mathbf{x}^{\boldsymbol{\gamma}'}=\mathbf{x}^{\boldsymbol{\beta}}\mathbf{x}^{\boldsymbol{\gamma}}$, $\{\mathbf{x}^{\boldsymbol{\beta}},\mathbf{x}^{\boldsymbol{\gamma}}\}\in E\Rightarrow\{\mathbf{x}^{\boldsymbol{\beta}'},\mathbf{x}^{\boldsymbol{\gamma}'}\}\in E$



- Iteratively perform support extension and chordal extension:

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

The moment-SOS hierarchy based on term sparsity

- The maximal cliques of $G_j^{(s)}$: $C_{j,1}^{(s)}, \ldots, C_{j,t_{j,s}}^{(s)}$
- TSSOS hierarchy:

$$\theta_r^{(s)} := \begin{cases} \inf & L_{\mathbf{y}}(f) \\ \text{s.t.} & M_{C_{0,i}^{(s)}}(\mathbf{y}) \succeq 0, \quad i = 1, \ldots, t_{0,s}, \\ & M_{C_{j,i}^{(s)}}(g_j\mathbf{y}) \succeq 0, \quad i = 1, \ldots, t_{j,s}, j = 1, \ldots, m, \\ & y_{\mathbf{0}} = 1. \end{cases}$$

The moment-SOS hierarchy based on term sparsity

- The maximal cliques of $G_j^{(s)}$: $C_{j,1}^{(s)}, \ldots, C_{j,t_{j,s}}^{(s)}$
- TSSOS hierarchy:

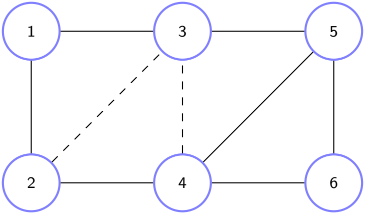
$$\theta_r^{(s)} := \begin{cases} \inf & L_{\mathbf{y}}(f) \\ \text{s.t.} & M_{C_{0,i}^{(s)}}(\mathbf{y}) \succeq 0, \quad i = 1, \ldots, t_{0,s}, \\ & M_{C_{j,i}^{(s)}}(g_j \mathbf{y}) \succeq 0, \quad i = 1, \ldots, t_{j,s}, j = 1, \ldots, m, \\ & y_{\mathbf{0}} = 1. \end{cases}$$

A two-level hierarchy of lower bounds

$$
\begin{array}{ccccccc}
\theta_{\underline{r}}^{(1)} & \leq & \theta_{\underline{r}}^{(2)} & \leq & \cdots & \leq & \theta_{\underline{r}} \\
\rotatebox{-90}{$\leq$} & & \rotatebox{-90}{$\leq$} & & & & \rotatebox{-90}{$\leq$} \\
\theta_{\underline{r}+1}^{(1)} & \leq & \theta_{\underline{r}+1}^{(2)} & \leq & \cdots & \leq & \theta_{\underline{r}+1} \\
\rotatebox{-90}{$\leq$} & & \rotatebox{-90}{$\leq$} & & & & \rotatebox{-90}{$\leq$} \\
\vdots & & \vdots & & \vdots & & \vdots \\
\rotatebox{-90}{$\leq$} & & \rotatebox{-90}{$\leq$} & & & & \rotatebox{-90}{$\leq$} \\
\theta_{r}^{(1)} & \leq & \theta_{r}^{(2)} & \leq & \cdots & \leq & \theta_{r} \\
\rotatebox{-90}{$\leq$} & & \rotatebox{-90}{$\leq$} & & & & \rotatebox{-90}{$\leq$} \\
\vdots & & \vdots & & \vdots & & \vdots
\end{array}
$$

Different choices of chordal extensions

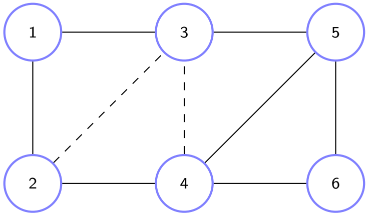
- chordal extension:
  - maximal chordal extension
  - (approximately) smallest chordal extension



- Fixing a relaxation order r, $(\theta_r^{(s)})_{s\geq 1}$ converges to θ_r in finitely many steps if the maximal chordal extension is chosen.

Different choices of chordal extensions

- chordal extension:
  - ➤ maximal chordal extension
  - ➤ (approximately) smallest chordal extension



- Fixing a relaxation order r, $(\theta_r^{(s)})_{s\geq 1}$ converges to θ_r in finitely many steps if the maximal chordal extension is chosen.

The connection to sign symmetries

- Sign symmetries: $f = x^4y^2 + 2x^2y^4 + xy + 1$, $f(x, y) = f(-x, -y)$
- The sign symmetries induce a partition of monomial bases:

$\mathbf{x}^{\alpha}, \mathbf{x}^{\beta}$ belong to the same block $\iff$ $\mathbf{x}^{\alpha+\beta}$ is invariant under the sign symmetries

Theorem (Wang, Magron, and Lasserre, 2021)

Fix a relaxation order r and assume the maximal chordal extension is chosen. The block structures arising from the TSSOS hierarchy converge to the one induced by the sign symmetries of the system.

The connection to sign symmetries

- Sign symmetries: $f = x^4y^2 + 2x^2y^4 + xy + 1$, $f(x, y) = f(-x, -y)$
- The sign symmetries induce a partition of monomial bases:

$\mathbf{x}^{\boldsymbol{\alpha}}, \mathbf{x}^{\boldsymbol{\beta}}$ belong to the same block $\iff$ $\mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}$ is invariant under the sign symmetries

Theorem (Wang, Magron, and Lasserre, 2021)

Fix a relaxation order r and assume the maximal chordal extension is chosen. The block structures arising from the TSSOS hierarchy converge to the one induced by the sign symmetries of the system.

The connection to sign symmetries

- Sign symmetries: $f = x^4y^2 + 2x^2y^4 + xy + 1$, $f(x, y) = f(-x, -y)$
- The sign symmetries induce a partition of monomial bases:

$\mathbf{x}^{\boldsymbol{\alpha}}, \mathbf{x}^{\boldsymbol{\beta}}$ belong to the same block $\iff$ $\mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}$ is invariant under the sign symmetries

Theorem (Wang, Magron, and Lasserre, 2021)

Fix a relaxation order r and assume the maximal chordal extension is chosen. The block structures arising from the TSSOS hierarchy converge to the one induced by the sign symmetries of the system.

Combining correlative sparsity with term sparsity

- The CS-TSSOS hierarchy:

1. Decomposing the variables with respect to the maximal cliques of the csp graph;
2. For each subsystem involving variables from one maximal clique, applying the above iterative procedure to exploit term sparsity.

Combining correlative sparsity with term sparsity

- The CS-TSSOS hierarchy:
  1. Decomposing the variables with respect to the maximal cliques of the csp graph;
  2. For each subsystem involving variables from one maximal clique, applying the above iterative procedure to exploit term sparsity.

SONC decomposition and second order cone representation

The second order cone representation of SONC cones

- Second order cone: $\mathbb{S}_+^2 := \{(a, b, c) \in \mathbb{R}^3 \mid \begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0\}$
- SONC cone: Given $\mathscr{A}, \mathscr{B}_1 \subseteq (2\mathbb{N})^n$ and $\mathscr{B}_2 \subseteq \mathbb{N}^n \backslash (2\mathbb{N})^n$,

$$\mathrm{SONC}_{\mathscr{A},\mathscr{B}_1,\mathscr{B}_2} := \{(\mathbf{c}_{\mathscr{A}}, \mathbf{d}_{\mathscr{B}_1}, \mathbf{d}_{\mathscr{B}_2}) \in \mathbb{R}_+^{|\mathscr{A}|} \times \mathbb{R}_+^{|\mathscr{B}_1|} \times \mathbb{R}^{|\mathscr{B}_2|} \mid \sum_{\alpha \in \mathscr{A}} c_\alpha \mathbf{x}^\alpha - \sum_{\beta \in \mathscr{B}_1 \cup \mathscr{B}_2} d_\beta \mathbf{x}^\beta \in \mathrm{SONC}\}$$

Theorem (Wang and Magron, 2020)

Any SONC cone admits a second order cone representation.

The second order cone representation of SONC cones

- Second order cone: $\mathbb{S}_+^2 := \{(a,b,c) \in \mathbb{R}^3 \mid \begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0\}$
- SONC cone: Given $\mathscr{A}, \mathscr{B}_1 \subseteq (2\mathbb{N})^n$ and $\mathscr{B}_2 \subseteq \mathbb{N}^n \backslash (2\mathbb{N})^n$,

$$\mathrm{SONC}_{\mathscr{A},\mathscr{B}_1,\mathscr{B}_2} := \{(\mathbf{c}_{\mathscr{A}}, \mathbf{d}_{\mathscr{B}_1}, \mathbf{d}_{\mathscr{B}_2}) \in \mathbb{R}_+^{|\mathscr{A}|} \times \mathbb{R}_+^{|\mathscr{B}_1|} \times \mathbb{R}^{|\mathscr{B}_2|} \mid \sum_{\boldsymbol{\alpha} \in \mathscr{A}} c_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}} - \sum_{\boldsymbol{\beta} \in \mathscr{B}_1 \cup \mathscr{B}_2} d_{\boldsymbol{\beta}} \mathbf{x}^{\boldsymbol{\beta}} \in \mathrm{SONC}\}$$

Theorem (Wang and Magron, 2020)

Any SONC cone admits a second order cone representation.

The second order cone representation of SONC cones

- Second order cone: $\mathbb{S}_+^2 := \{(a, b, c) \in \mathbb{R}^3 \mid \begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0\}$
- SONC cone: Given $\mathscr{A}, \mathscr{B}_1 \subseteq (2\mathbb{N})^n$ and $\mathscr{B}_2 \subseteq \mathbb{N}^n \backslash (2\mathbb{N})^n$,

$$\mathrm{SONC}_{\mathscr{A},\mathscr{B}_1,\mathscr{B}_2} := \{(\mathbf{c}_{\mathscr{A}}, \mathbf{d}_{\mathscr{B}_1}, \mathbf{d}_{\mathscr{B}_2}) \in \mathbb{R}_+^{|\mathscr{A}|} \times \mathbb{R}_+^{|\mathscr{B}_1|} \times \mathbb{R}^{|\mathscr{B}_2|} \mid \sum_{\boldsymbol{\alpha} \in \mathscr{A}} c_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}} - \sum_{\boldsymbol{\beta} \in \mathscr{B}_1 \cup \mathscr{B}_2} d_{\boldsymbol{\beta}} \mathbf{x}^{\boldsymbol{\beta}} \in \mathrm{SONC}\}$$

Theorem (Wang and Magron, 2020)

Any SONC cone admits a second order cone representation.

Circuit polynomial and geometric mean inequality

- $f = \sum_{\boldsymbol{\alpha}\in\mathscr{A}} c_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}} - d_{\boldsymbol{\beta}}\mathbf{x}^{\boldsymbol{\beta}}$ ($d_{\boldsymbol{\beta}} > 0$) a circuit polynomial
- There exist unique barycentric coordinates $(\lambda_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathscr{A}} \subseteq \mathbb{R}^{+}$ s.t.

$$\sum_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}} = 1, \text{ and } \boldsymbol{\beta} = \sum_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}}\boldsymbol{\alpha}$$

- Then,

$$f \geq 0 \iff \prod_{\boldsymbol{\alpha}\in\mathscr{A}} \left(\frac{c_{\boldsymbol{\alpha}}}{\lambda_{\boldsymbol{\alpha}}}\right)^{\lambda_{\boldsymbol{\alpha}}} \geq d_{\boldsymbol{\beta}} \iff \prod_{\boldsymbol{\alpha}\in\mathscr{A}} c_{\boldsymbol{\alpha}}^{\lambda_{\boldsymbol{\alpha}}} \geq \prod_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}}^{\lambda_{\boldsymbol{\alpha}}} d_{\boldsymbol{\beta}}$$

Circuit polynomial and geometric mean inequality

- $f = \sum_{\boldsymbol{\alpha}\in\mathscr{A}} c_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}} - d_{\boldsymbol{\beta}}\mathbf{x}^{\boldsymbol{\beta}}$ ($d_{\boldsymbol{\beta}} > 0$) a circuit polynomial
- There exist unique barycentric coordinates $(\lambda_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathscr{A}} \subseteq \mathbb{R}^+$ s.t.

$$\sum_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}} = 1, \text{ and } \boldsymbol{\beta} = \sum_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}}\boldsymbol{\alpha}$$

- Then,

$$f \geq 0 \iff \prod_{\boldsymbol{\alpha}\in\mathscr{A}} \left(\frac{c_{\boldsymbol{\alpha}}}{\lambda_{\boldsymbol{\alpha}}}\right)^{\lambda_{\boldsymbol{\alpha}}} \geq d_{\boldsymbol{\beta}} \iff \prod_{\boldsymbol{\alpha}\in\mathscr{A}} c_{\boldsymbol{\alpha}}^{\lambda_{\boldsymbol{\alpha}}} \geq \prod_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}}^{\lambda_{\boldsymbol{\alpha}}} d_{\boldsymbol{\beta}}$$

Circuit polynomial and geometric mean inequality

- $f = \sum_{\boldsymbol{\alpha}\in\mathscr{A}} c_{\boldsymbol{\alpha}}\mathbf{x}^{\boldsymbol{\alpha}} - d_{\boldsymbol{\beta}}\mathbf{x}^{\boldsymbol{\beta}}$ ($d_{\boldsymbol{\beta}} > 0$) a circuit polynomial
- There exist unique barycentric coordinates $(\lambda_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\mathscr{A}} \subseteq \mathbb{R}^+$ s.t.

$$\sum_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}} = 1, \text{ and } \boldsymbol{\beta} = \sum_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}}\boldsymbol{\alpha}$$

- Then,

$$f \geq 0 \iff \prod_{\boldsymbol{\alpha}\in\mathscr{A}} \left(\frac{c_{\boldsymbol{\alpha}}}{\lambda_{\boldsymbol{\alpha}}}\right)^{\lambda_{\boldsymbol{\alpha}}} \geq d_{\boldsymbol{\beta}} \iff \prod_{\boldsymbol{\alpha}\in\mathscr{A}} c_{\boldsymbol{\alpha}}^{\lambda_{\boldsymbol{\alpha}}} \geq \prod_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}}^{\lambda_{\boldsymbol{\alpha}}} d_{\boldsymbol{\beta}}$$

Mediated sets and sum of binomial squares

- $\mathscr{A}$-mediated set: $M \subseteq \mathbb{Q}^n$, for each $\mathbf{w} \in M$,

$$\exists \mathbf{u} \neq \boldsymbol{v} \in M \cup \mathscr{A}, \text{ s.t. } \mathbf{w} = \frac{1}{2}(\mathbf{u} + \boldsymbol{v})$$

Theorem (Wang and Magron, 2020)

A circuit polynomial $f = \sum_{\alpha \in \mathscr{A}} c_\alpha \mathbf{x}^\alpha - d_\beta \mathbf{x}^\beta \in \mathbb{R}[\mathbf{x}]$ ($d_\beta > 0$) is nonnegative if and only if $f = \sum_{\mathbf{w}_i \in M} (a_i \mathbf{x}^{\frac{1}{2}\mathbf{u}_i} - b_i \mathbf{x}^{\frac{1}{2}\boldsymbol{v}_i})^2$, $a_i, b_i \in \mathbb{R}$ for any $\mathscr{A}$-mediated set M containing β.

Mediated sets and sum of binomial squares

- $\mathscr{A}$-mediated set: $M \subseteq \mathbb{Q}^n$, for each $\mathbf{w} \in M$,

$$\exists \mathbf{u} \neq \boldsymbol{v} \in M \cup \mathscr{A}, \text{ s.t. } \mathbf{w} = \frac{1}{2}(\mathbf{u} + \boldsymbol{v})$$

Theorem (Wang and Magron, 2020)

A circuit polynomial $f = \sum_{\boldsymbol{\alpha} \in \mathscr{A}} c_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}} - d_{\boldsymbol{\beta}} \mathbf{x}^{\boldsymbol{\beta}} \in \mathbb{R}[\mathbf{x}]$ ($d_{\boldsymbol{\beta}} > 0$) is nonnegative if and only if $f = \sum_{\mathbf{w}_i \in M} (a_i \mathbf{x}^{\frac{1}{2}\mathbf{u}_i} - b_i \mathbf{x}^{\frac{1}{2}\boldsymbol{v}_i})^2$, $a_i, b_i \in \mathbb{R}$ for any $\mathscr{A}$-mediated set M containing $\boldsymbol{\beta}$.

Correspondences

construct a second order cone representation for a SONC cone $\mathcal{S}$

$$\Updownarrow$$

rewrite any circuit polynomial $f \in \mathcal{S}$ as a sum of binomial squares

$$\Updownarrow$$

for any lattice point $\boldsymbol{\beta}$, find an $\mathscr{A}$-mediated set M containing $\boldsymbol{\beta}$

$$\Updownarrow$$

rewrite $\prod_{\boldsymbol{\alpha}\in\mathscr{A}} c_{\boldsymbol{\alpha}}^{\lambda_{\boldsymbol{\alpha}}} \geq \prod_{\boldsymbol{\alpha}\in\mathscr{A}} \lambda_{\boldsymbol{\alpha}}^{\lambda_{\boldsymbol{\alpha}}} d_{\boldsymbol{\beta}}$ as a set of quadratic inequalities

Optimal second order cone representations for geometric mean inequalities

Question

Given an inequality $x_1^{s_1} \cdots x_m^{s_m} \geq x_{m+1}^{\sum_{i=1}^m s_i}$ $(x_i \in \mathbb{R}^+, s_i \in \mathbb{N}^*)$, construct an equivalent representation using as few quadratic inequalities as possible.

Example: $x_1^3 x_2^8 \geq x_3^{11} \iff x_2x_4 \geq x_3^2, x_5x_6 \geq x_4^2, x_1x_3 \geq x_5^2, x_3x_5 \geq x_6^2$

- $L(s_1, \ldots, s_m)$: minimum number of quadratic inequalities
  - $L(s_1, s_2) = \lceil \log_2 (s_1 + s_2) \rceil$
  - $L(s_1, \ldots, s_m) \geq \lceil \log_2 (\sum_{i=1}^m s_i) \rceil$

Optimal second order cone representations for geometric mean inequalities

Question

Given an inequality $x_1^{s_1}\cdots x_m^{s_m}\ge x_{m+1}^{\sum_{i=1}^m s_i}$ ($x_i\in\mathbb{R}^+, s_i\in\mathbb{N}^*$), construct an equivalent representation using as few quadratic inequalities as possible.

Example: $x_1^3x_2^8\ge x_3^{11} \iff x_2x_4\ge x_3^2, x_5x_6\ge x_4^2, x_1x_3\ge x_5^2, x_3x_5\ge x_6^2$

- $L(s_1,\ldots,s_m)$: minimum number of quadratic inequalities
  - $L(s_1,s_2)=\lceil\log_2(s_1+s_2)\rceil$
  - $L(s_1,\ldots,s_m)\ge\lceil\log_2(\sum_{i=1}^m s_i)\rceil$

Optimal second order cone representations for geometric mean inequalities

Question

Given an inequality $x_1^{s_1} \cdots x_m^{s_m} \geq x_{m+1}^{\sum_{i=1}^m s_i}$ $(x_i \in \mathbb{R}^+, s_i \in \mathbb{N}^*)$, construct an equivalent representation using as few quadratic inequalities as possible.

Example: $x_1^3 x_2^8 \geq x_3^{11} \iff x_2 x_4 \geq x_3^2, x_5 x_6 \geq x_4^2, x_1 x_3 \geq x_5^2, x_3 x_5 \geq x_6^2$

- $L(s_1, \ldots, s_m)$: minimum number of quadratic inequalities
  - $L(s_1, s_2) = \lceil \log_2 (s_1 + s_2) \rceil$
  - $L(s_1, \ldots, s_m) \geq \lceil \log_2 (\sum_{i=1}^m s_i) \rceil$

A heuristic algorithm for (nearly) optimal second order cone representations

$$x_1^{s_1}\cdots x_m^{s_m}\ge x_{m+1}^{\sum_{i=1}^m s_i}$$

$$\Updownarrow$$

$$x_1^{s_1}\cdots x_i^{s_i-s_i}\cdots x_j^{s_j-s_i}\cdots x_m^{s_m}x_{m+2}^{2s_i}\ge x_{m+1}^{\sum_{i=1}^m s_i},\ x_ix_j\ge x_{m+2}^2$$

$$\Updownarrow$$

$$\vdots$$

$$x_{i_1}x_{j_1}\ge x_{m+1}^2,\dots,x_{i_n}x_{j_n}\ge x_{m+n}^2,\ i_k,j_k\in\{1,2,\dots,m+n\}$$

Software and numerical experiments

- TSSOS:

https://github.com/wangjie212/TSSOS

- SONCSOCP:

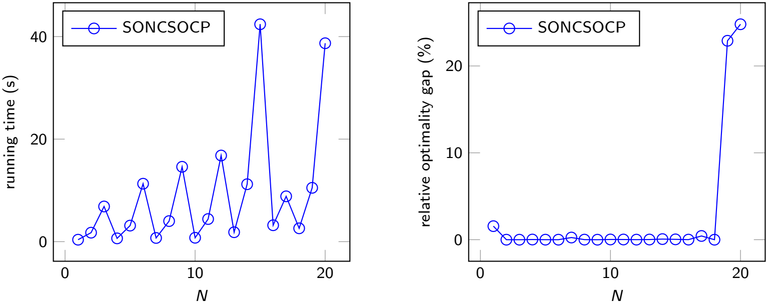
https://github.com/wangjie212/SONCSOCP

Optimal power flow (AC-OPF)

n	m+m'	CS ($r = 2$)				CS+TS ($r = 2, s = 1$)			
		mb	opt	time (s)	gap	mb	opt	time (s)	gap
12	28	28	1.1242e4	0.21	0.00%	22	1.1242e4	0.09	0.00%
20	55	28	1.7543e4	0.56	0.05%	22	1.7543e4	0.30	0.05%
72	297	45	4.9927e3	4.43	0.07%	22	4.9920e3	2.69	0.08%
114	315	120	7.6943e4	94.9	0.00%	39	7.6942e4	14.8	0.00%
344	1325	253	—	—	—	73	1.0470e5	169	0.50%
348	1809	253	—	—	—	34	1.2096e5	201	0.03%
766	3322	153	3.3072e6	585	0.68%	44	3.3042e6	33.9	0.77%
1112	4613	496	—	—	—	31	7.2396e4	410	0.25%
4356	18257	378	—	—	—	27	1.3953e6	934	0.51%
6698	29283	1326	—	—	—	76	5.9858e5	1886	0.47%

Optimization with polynomials of high degree

number of variables: 10, degree: $20 \sim 50$, number of terms: $30 \sim 300$



Summary

- Term sparsity opens a new window for exploiting sparsity in polynomial optimization
- When appropriate sparsity patterns are accessible, significantly improve the scalability of the moment-SOS hierarchy
- SONC decompositions work for sparse polynomials of high degree
- Further investigations on $L(s_1, \ldots, s_m)$

Summary

- Term sparsity opens a new window for exploiting sparsity in polynomial optimization
- When appropriate sparsity patterns are accessible, significantly improve the scalability of the moment-SOS hierarchy
- SONC decompositions work for sparse polynomials of high degree
- Further investigations on $L(s_1, \ldots, s_m)$

Summary

- Term sparsity opens a new window for exploiting sparsity in polynomial optimization
- When appropriate sparsity patterns are accessible, significantly improve the scalability of the moment-SOS hierarchy
- SONC decompositions work for sparse polynomials of high degree
- Further investigations on $L(s_1, \ldots, s_m)$

Summary

- Term sparsity opens a new window for exploiting sparsity in polynomial optimization
- When appropriate sparsity patterns are accessible, significantly improve the scalability of the moment-SOS hierarchy
- SONC decompositions work for sparse polynomials of high degree
- Further investigations on $L(s_1, \ldots, s_m)$

Main references

- **Jie Wang**, *Nonnegative Polynomials and Circuit Polynomials*, SIAM Journal on Applied Algebra and Geometry, 2022.
- **Jie Wang**, Victor Magron and Jean B. Lasserre, *TSSOS: A Moment-SOS hierarchy that exploits term sparsity*, SIAM Journal on Optimization, 2021.
- **Jie Wang**, Victor Magron and Jean B. Lasserre, *Chordal-TSSOS: a moment-SOS hierarchy that exploits term sparsity with chordal extension*, SIAM Journal on Optimization, 2021.
- **Jie Wang**, Victor Magron, Jean B. Lasserre and Ngoc H. A. Mai, *CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization*, arXiv:2005.02828, 2020.

Thanks for your attention!

https://wangjie212.github.io/jiewang