Structured Polynomial Optimization: A Unified Approach For Global Optimization

Jie Wang

Academy of Mathematics and Systems Science, CAS

July 20, 2023

Collaborators

Jean B. Lasserre

Victor Magron

Outline

Outline

1. Polynomial optimization and the moment-SOS hierarchy
2. Improve scalability by exploiting algebraic structures
3. Numerical experiments and applications

Outline

1. Polynomial optimization and the moment-SOS hierarchy

2. Improve scalability by exploiting algebraic structures

3. Numerical experiments and applications

Polynomial optimization

- Polynomial optimization problem (POP):

$$f_{\min} := \begin{cases} \inf\limits_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \geq 0, \quad i = 1, \ldots, m \end{cases}$$

- non-convex, NP-hard
- optimal power flow, computer vision, combinatorial optimization, neutral networks, signal processing, quantum information...

Polynomial optimization

- Polynomial optimization problem (POP):

$$f_{\min} := \begin{cases} \inf\limits_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \geq 0, \quad i = 1, \ldots, m \end{cases}$$

- non-convex, NP-hard
- optimal power flow, computer vision, combinatorial optimization, neutral networks, signal processing, quantum information...

Polynomial optimization

- Polynomial optimization problem (POP):

$$f_{\min} := \begin{cases} \inf\limits_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \geq 0, \quad i = 1, \ldots, m \end{cases}$$

- non-convex, NP-hard

- optimal power flow, computer vision, combinatorial optimization, neutral networks, signal processing, quantum information...

Why polynomial optimization?

- closely related to real algebraic geometry: the theory of positive polynomials, convex algebraic geometry
- be able to compute the globally optimal value/solutions: the Moment-SOS hierarchy
- closely related to theoretical computer science: the theory of approximation algorithms, the theory of complexity
- Powerful modelling ability: QCQP, binary program, (mixed) integer (non-)linear program and so on

Why polynomial optimization?

- closely related to real algebraic geometry: the theory of positive polynomials, convex algebraic geometry
- be able to compute the globally optimal value/solutions: the Moment-SOS hierarchy
- closely related to theoretical computer science: the theory of approximation algorithms, the theory of complexity
- Powerful modelling ability: QCQP, binary program, (mixed) integer (non-)linear program and so on

Why polynomial optimization?

- closely related to real algebraic geometry: the theory of positive polynomials, convex algebraic geometry
- be able to compute the globally optimal value/solutions: the Moment-SOS hierarchy
- closely related to theoretical computer science: the theory of approximation algorithms, the theory of complexity
- Powerful modelling ability: QCQP, binary program, (mixed) integer (non-)linear program and so on

Why polynomial optimization?

- closely related to real algebraic geometry: the theory of positive polynomials, convex algebraic geometry
- be able to compute the globally optimal value/solutions: the Moment-SOS hierarchy
- closely related to theoretical computer science: the theory of approximation algorithms, the theory of complexity
- Powerful modelling ability: QCQP, binary program, (mixed) integer (non-)linear program and so on

Continuous optimization appears as POPs

Continuous convex and nonconvex optimization problems with linear and/or quadratic costs and constraints:

$$\begin{cases} \inf\limits_{\mathbf{x}\in\mathbb{R}^n} & \mathbf{x}^\intercal A_0\mathbf{x} + b_0^\intercal\mathbf{x} \\ \text{s.t.} & \mathbf{x}^\intercal A_i\mathbf{x} + b_i^\intercal\mathbf{x} - c_i \geq 0, \quad i = 1,\ldots,m. \end{cases}$$

Continuous optimization approximated by POPs

- Any continuous function can be approximated by polynomials as closely as desired.
- Any continuous optimization problem can be approximated by POPs as closely as desired.

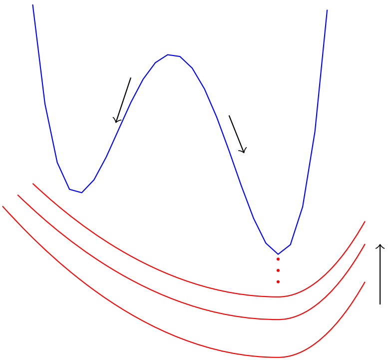
Continuous optimization approximated by POPs

- Any continuous function can be approximated by polynomials as closely as desired.
- Any continuous optimization problem can be approximated by POPs as closely as desired.

Discrete optimization

- ± 1 variables: $x \in \{-1, +1\} \iff x^2 - 1 = 0$
- $0/1$ variables: $x \in \{0, 1\} \iff x(x-1) = 0$
- integer variables: $x \in \{1, 2, \ldots, t\} \iff (x-1)(x-2)\cdots(x-t) = 0$

Non-convexity of polynomial optimization



Example (moment relaxation)

$$\begin{cases} \inf\limits_{\mathbf{x}} & x_1^2 + x_1x_2 + x_2^2 \\ \text{s.t.} & 1 - x_1^2 \geq 0, 1 - x_2^2 \geq 0 \end{cases} \Longleftrightarrow \begin{cases} \inf\limits_{\mathbf{x}} & x_1^2 + x_1x_2 + x_2^2 \\ \text{s.t.} & \begin{bmatrix} 1 & x_1 & x_2 \\ x_1 & x_1^2 & x_1x_2 \\ x_2 & x_1x_2 & x_2^2 \end{bmatrix} = [1, x_1, x_2] \cdot [1, x_1, x_2]^\intercal \succeq 0, \\ & 1 - x_1^2 \geq 0, 1 - x_2^2 \geq 0 \end{cases}$$

$$\Longleftrightarrow \begin{cases} \inf\limits_{\mathbf{y}} & y_{2,0} + y_{1,1} + y_{0,2} \\ \text{s.t.} & \begin{bmatrix} 1 & y_{1,0} & y_{0,1} \\ y_{1,0} & y_{2,0} & y_{1,1} \\ y_{0,1} & y_{1,1} & y_{0,2} \end{bmatrix} \succeq 0, \\ & 1 - y_{2,0} \geq 0, 1 - y_{0,2} \geq 0, \\ & \exists \mathbf{x} \in \mathbb{R}^2 \text{ s.t. } \mathbf{y} = (\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_1^2, \mathrm{x}_1\mathrm{x}_2, \mathrm{x}_2^2) \end{cases} \xRightarrow{\textit{relax}} \begin{cases} \inf\limits_{\mathbf{y}} & y_{2,0} + y_{1,1} + y_{0,2} \\ \text{s.t.} & \begin{bmatrix} 1 & y_{1,0} & y_{0,1} \\ y_{1,0} & y_{2,0} & y_{1,1} \\ y_{0,1} & y_{1,1} & y_{0,2} \end{bmatrix} \succeq 0, \\ & 1 - y_{2,0} \geq 0, 1 - y_{0,2} \geq 0 \end{cases}$$

The hierarchy of moment relaxations

- The hierarchy of moment relaxations (Lasserre 2001):

$$\theta_r := \begin{cases} \inf\limits_{\mathbf{y}} & L_{\mathbf{y}}(f) \\ \text{s.t.} & \mathbf{M}_r(\mathbf{y}) \succeq 0, \\ & \mathbf{M}_{r-d_i}(g_i\mathbf{y}) \succeq 0, \quad i = 1, \ldots, m, \\ & y_{\mathbf{0}} = 1. \end{cases}$$

Example (dual SOS relaxation)

$$\begin{cases} \inf_{\mathbf{x}} & x_1^2 + x_1x_2 + x_2^2 \\ \text{s.t.} & 1 - x_1^2 \geq 0, 1 - x_2^2 \geq 0 \end{cases} \iff \begin{cases} \sup_{\lambda} & \lambda \\ \text{s.t.} & x_1^2 + x_1x_2 + x_2^2 - \lambda \geq 0, \forall \mathbf{x} \in \mathbb{R}^2 \text{ s.t. } (1 - x_1^2 \geq 0, 1 - x_2^2 \geq 0) \end{cases}$$

$$\xRightarrow{\text{strengthen}} \begin{cases} \sup_{\lambda,\sigma_i} & \lambda \\ \text{s.t.} & x_1^2 + x_1x_2 + x_2^2 - \lambda = \sigma_0 + \sigma_1(1 - x_1^2) + \sigma_2(1 - x_2^2), \\ & \sigma_0, \sigma_1, \sigma_2 \in \text{SOS} \end{cases}$$

The hierarchy of dual SOS relaxations

- The hierarchy of dual SOS relaxations (Parrilo 2000 & Lasserre 2001):

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

The Moment-SOS/Lasserre's hierarchy

$$
\begin{array}{ccccccc}
 & & & f_{\min} & & & \\
 & & \rotatebox{45}{$\geq$} & & \rotatebox{-45}{$\leq$} & & \\
 & \vdots & & & & \vdots & \\
 & \geq & & & & \geq & \\
\text{(Moment relaxation)} & \theta_r & & \text{“} = \text{”} & & \theta_r^* & \text{(dual SOS relaxation)} \\
 & \geq & & & & \geq & \\
 & \vdots & & & & \vdots & \\
 & \geq & & & & \geq & \\
 & \theta_{r_{\min}} & & \text{“} = \text{”} & & \theta_{r_{\min}}^* &
\end{array}
$$

Asymptotical convergence and finite convergence

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

➤ $\theta_r \nearrow f_{\min}$ and $\theta_r^* \nearrow f_{\min}$ as $r \to \infty$ (**Putinar's Positivstellensatz, 1993**)

➤ Finite convergence happens generically (**Nie, 2014**)

Asymptotical convergence and finite convergence

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

➤ $\theta_r \nearrow f_{\min}$ and $\theta_r^* \nearrow f_{\min}$ as $r \to \infty$ (**Putinar's Positivstellensatz, 1993**)

➤ Finite convergence happens generically (**Nie, 2014**)

Asymptotical convergence and finite convergence

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

➤ $\theta_r \nearrow f_{\min}$ and $\theta_r^* \nearrow f_{\min}$ as $r \to \infty$ (**Putinar's Positivstellensatz, 1993**)

➤ Finite convergence happens generically (**Nie, 2014**)

Detecting global optimality

- The moment relaxation achieves global optimality ($\theta_r = f_{\min}$) when one of the following conditions holds:
  - (**flat extension**) For $r_0 \le r' \le r$, $\operatorname{rank} \mathbf{M}_{r'-r_0}(\mathbf{y}) = \operatorname{rank} \mathbf{M}_{r'}(\mathbf{y})$
    - $\rightsquigarrow$ Extract $\operatorname{rank} \mathbf{M}_{r'}(\mathbf{y})$ globally optimal solutions
  - $\operatorname{rank} \mathbf{M}_{r_{\min}}(\mathbf{y}) = 1$
    - $\rightsquigarrow$ Extract one globally optimal solution

Detecting global optimality

- The moment relaxation achieves global optimality ($\theta_r = f_{\min}$) when one of the following conditions holds:
  - (**flat extension**) For $r_0 \leq r' \leq r$, $\mathrm{rank}\, \mathbf{M}_{r'-r_0}(\mathbf{y}) = \mathrm{rank}\, \mathbf{M}_{r'}(\mathbf{y})$
    - $\rightsquigarrow$ Extract $\mathrm{rank}\, \mathbf{M}_{r'}(\mathbf{y})$ globally optimal solutions
  - $\mathrm{rank}\, \mathbf{M}_{r_{\min}}(\mathbf{y}) = 1$
    - $\rightsquigarrow$ Extract one globally optimal solution

Detecting global optimality

- The moment relaxation achieves global optimality ($\theta_r = f_{\min}$) when one of the following conditions holds:
  - (**flat extension**) For $r_0 \leq r' \leq r$, $\operatorname{rank} \mathbf{M}_{r'-r_0}(\mathbf{y}) = \operatorname{rank} \mathbf{M}_{r'}(\mathbf{y})$

    $\rightsquigarrow$ Extract $\operatorname{rank} \mathbf{M}_{r'}(\mathbf{y})$ globally optimal solutions
  - $\operatorname{rank} \mathbf{M}_{r_{\min}}(\mathbf{y}) = 1$

    $\rightsquigarrow$ Extract one globally optimal solution

Extension – complex polynomial optimization

- Complex polynomial optimization problem (CPOP):

$$\begin{cases} \inf\limits_{\mathbf{z}\in\mathbb{C}^n} & f(\mathbf{z},\overline{\mathbf{z}}) \\ \text{s.t.} & g_i(\mathbf{z},\overline{\mathbf{z}}) \geq 0, \quad i=1,\ldots,m, \\ & h_j(\mathbf{z},\overline{\mathbf{z}}) = 0, \quad j=1,\ldots,l. \end{cases}$$

$\rightsquigarrow$ The moment-HSOS hierarchy

$\rightsquigarrow$ optimal power flow

Extension – trigonometric polynomial optimization

- Trigonometric polynomial optimization problem:

$$\begin{cases} \inf\limits_{x\in[0,2\pi)^n} & f(\sin x_1,\ldots,\sin x_n,\cos x_1,\ldots,\cos x_n) \\ \text{s.t.} & g_i(\sin x_1,\ldots,\sin x_n,\cos x_1,\ldots,\cos x_n)\ge 0, \quad i=1,\ldots,m, \\ & h_j(\sin x_1,\ldots,\sin x_n,\cos x_1,\ldots,\cos x_n)=0, \quad j=1,\ldots,l. \end{cases}$$

$\rightsquigarrow$ The moment-HSOS hierarchy

$\rightsquigarrow$ sigal processing

Extension – noncommutative polynomial optimization

- Eigenvalue optimization problem:

$$\begin{cases} \inf\limits_{X} & \text{eig}\, f(X) = f(X_1, \ldots, X_n) \\ \text{s.t.} & g_i(X) \geq 0, \quad i = 1, \ldots, m, \\ & h_j(X) = 0, \quad j = 1, \ldots, l. \end{cases}$$

⇝ The moment-SOHS hierarchy

⇝ linear Bell inequality

Extension – noncommutative polynomial optimization

- Trace optimization problem:

$$\begin{cases} \inf\limits_{X} & \operatorname{tr} f(X) = f(X_1, \ldots, X_n) \\ \text{s.t.} & g_i(X) \geq 0, \quad i = 1, \ldots, m, \\ & h_j(X) = 0, \quad j = 1, \ldots, l. \end{cases}$$

$\rightsquigarrow$ The tracial moment-SOHS hierarchy

Extension – trace/state polynomial optimization

- trace polynomial: $\mathrm{tr}(x_1^2)x_2x_1 + \mathrm{tr}(x_1)\mathrm{tr}(x_2x_1x_2)$, $x_1, \ldots, x_n \in \mathcal{B}(\mathcal{H})$
- state polynomial: $\varsigma(x_1^2)x_2x_1 + \varsigma(x_1)\varsigma(x_2x_1x_2)$, $x_1, \ldots, x_n \in \mathcal{B}(\mathcal{H})$, ς is a formal state (i.e., a positive unital linear functional) on $\mathcal{B}(\mathcal{H})$

  $\rightsquigarrow$ The moment-SOHS hierarchy

  $\rightsquigarrow$ nonlinear Bell inequality

Extension – polynomial matrix optimization

- Robust polynomial matrix inequality optimization:

$$\begin{cases} \inf_{\mathbf{y}\in Y} & f(\mathbf{y}) \\ \text{s.t.} & P(\mathbf{y},\mathbf{x}) \succeq 0, \ \forall \mathbf{x} \in X. \end{cases}$$

$\rightsquigarrow$ The moment-SOS hierarchy

$\rightsquigarrow$ robust polynomial semidefinite program

Extension – polynomial dynamic system

- Polynomial dynamic system:

$$\begin{cases} \dot{x}_1 &= f_1(\mathbf{x}), \\ \dot{x}_2 &= f_2(\mathbf{x}), \\ & \vdots \\ \dot{x}_n &= f_n(\mathbf{x}), \end{cases}$$

$\rightsquigarrow$ The moment-SOS hierarchy

$\rightsquigarrow$ maximal (controlled) invariant set, attraction region, global attractor, reachable set, optimal control

The scalability issue of the moment-SOS hierarchy

- The size of SDP corresponding to the r-th SOS relaxation:

  1. PSD constraint: $\binom{n+r}{r}$
  2. #equality constraint: $\binom{n+2r}{2r}$

- $r = 2$, $n < 30$ (Mosek)
- Exploiting algebraic structures:
  - POP
  - SDP

The scalability issue of the moment-SOS hierarchy

- The size of SDP corresponding to the r-th SOS relaxation:
  1. PSD constraint: $\binom{n+r}{r}$
  2. #equality constraint: $\binom{n+2r}{2r}$
- $r = 2$, $n < 30$ (Mosek)
- Exploiting algebraic structures:
  - ➤ POP
  - ➤ SDP

The scalability issue of the moment-SOS hierarchy

- The size of SDP corresponding to the r-th SOS relaxation:
  1. PSD constraint: $\binom{n+r}{r}$
  2. #equality constraint: $\binom{n+2r}{2r}$
- $r = 2,\ n < 30$ (Mosek)
- Exploiting algebraic structures:
  - POP
  - SDP

Quotient ring

- Equality constraints: $h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, l$

- Build the moment-SOS hierarchy on the quotient ring

$$\mathbb{R}[\mathbf{x}]/(h_1(\mathbf{x}), \ldots, h_l(\mathbf{x}))$$

$\rightsquigarrow$ Gröbner basis

Quotient ring

- Equality constraints: $h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, l$

- Build the moment-SOS hierarchy on the quotient ring

$$\mathbb{R}[\mathbf{x}]/(h_1(\mathbf{x}), \ldots, h_l(\mathbf{x}))$$

$\rightsquigarrow$ Gröbner basis

Symmetry

- permutation symmetry: $(x_1, \ldots, x_n) \to (x_{\tau(1)}, \ldots, x_{\tau(n)})$
- translation symmetry: $(x_1, \ldots, x_n) \to (x_{1+i}, \ldots, x_{n+i})$, $x_{n+i} = x_i$
- sign symmetry: $(x_1, \ldots, x_n) \to (-x_1, \ldots, -x_n)$
- conjugate symmetry: $\mathbf{z} \to \bar{\mathbf{z}}$
- $\mathbb{T}$-symmetry: $\mathbf{z} \to e^{\mathbf{i}\theta}\mathbf{z}$

$\rightsquigarrow$ lead to block-diagonal moment-SOS hierarchies

Symmetry

- permutation symmetry: $(x_1, \dots, x_n) \to (x_{\tau(1)}, \dots, x_{\tau(n)})$
- translation symmetry: $(x_1, \dots, x_n) \to (x_{1+i}, \dots, x_{n+i})$, $x_{n+i} = x_i$
- sign symmetry: $(x_1, \dots, x_n) \to (-x_1, \dots, -x_n)$
- conjugate symmetry: $\mathbf{z} \to \bar{\mathbf{z}}$
- $\mathbb{T}$-symmetry: $\mathbf{z} \to e^{\mathbf{i}\theta}\mathbf{z}$

$\rightsquigarrow$ lead to block-diagonal moment-SOS hierarchies

Correlative sparsity (Waki et al. 2006)

- Correlative sparsity pattern 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 in the same constraint polynomial g_k
- For each maximal clique of $G^{\mathrm{csp}}(V,E)$, do

$$I_k \longmapsto \mathbf{M}_r(\mathbf{y}, I_k), \mathbf{M}_{r-d_i}(g_i\mathbf{y}, I_k)$$

Correlative sparsity (Waki et al. 2006)

- Correlative sparsity pattern 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 in the same

constraint polynomial g_k

- For each maximal clique of $G^{\text{csp}}(V, E)$, do

$$I_k \longmapsto \mathbf{M}_r(\mathbf{y}, I_k), \mathbf{M}_{r-d_i}(g_i\mathbf{y}, I_k)$$

Term sparsity (Wang & Magron & Lasserre, 2021)

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

$$
\begin{array}{c} \\ \vdots \\ \boldsymbol{\beta} \\ \vdots \end{array}
\begin{array}{c} \cdots \quad \boldsymbol{\alpha} \quad \cdots \\ \begin{bmatrix} & \vdots & \\ \cdots & y_{\boldsymbol{\alpha}+\boldsymbol{\beta}} & \cdots \\ & \vdots & \end{bmatrix} \end{array} = \mathbf{M}_r(\mathbf{y})
$$

Correlative-term sparsity

1. Decompose the whole set of variables into cliques by exploiting correlative sparsity
2. Exploit term sparsity for each subsystem

Correlative-term sparsity

1. Decompose the whole set of variables into cliques by exploiting correlative sparsity
2. Exploit term sparsity for each subsystem

Structures of the SOS problem

- Orthogonality: $\langle A_i, A_j \rangle = 0, \quad \forall i \neq j$

$$\begin{cases} \sup\limits_{X,x} & c^{\intercal} x \\ \text{s.t.} & \langle A_i, X \rangle + B_i x = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0 \end{cases}$$

Structures of the moment problem

- Low-rank: $\mathrm{rank}(\mathbf{M}^{\mathrm{opt}}) \ll n$
- Unital diagonal: $\mathrm{diag}(\mathbf{M}) = \mathbf{1}$
- Unital trace: $\mathrm{tr}(\mathbf{M}) = 1$

$$\begin{cases} \inf_{X \in \mathbb{R}^{n\times n}} & \langle C, X\rangle \\ \text{s.t.} & \langle A_i, X\rangle = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0 \end{cases}$$

Solving low-rank SDPs via manifold optimization

- Degenerate: $\geq$ 2nd relaxation $\rightsquigarrow$ $m \gg n$ **Challenging!**
- Low-rank: $\operatorname{rank} \mathbf{M}^{\text{opt}} \ll n \rightsquigarrow \mathbf{M} = YY^{\intercal}$, $Y \in \mathbb{R}^{n\times p}$ **Burer-Monteiro**
  - $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p}\}$
- Unital diagonal: $\operatorname{diag}(\mathbf{M}) = \mathbf{1}$
  - $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p} \mid \|Y(k,:)\| = 1, k = 1, \ldots, n\}$
- Unital trace: $\operatorname{tr}(\mathbf{M}) = 1$
  - $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p} \mid \|Y\|_F = 1\}$

Solving low-rank SDPs via manifold optimization

- Degenerate: $\geq$ 2nd relaxation $\rightsquigarrow m \gg n$ **Challenging!**
- Low-rank: $\mathrm{rank}\,\mathbf{M}^{\mathrm{opt}} \ll n \rightsquigarrow \mathbf{M} = YY^{\intercal}$, $Y \in \mathbb{R}^{n\times p}$ **Burer-Monteiro**
  - ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p}\}$
- Unital diagonal: $\mathrm{diag}(\mathbf{M}) = \mathbf{1}$
  - ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p} \mid \|Y(k,:)\| = 1, k = 1, \ldots, n\}$
- Unital trace: $\mathrm{tr}(\mathbf{M}) = 1$
  - ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p} \mid \|Y\|_F = 1\}$

Solving low-rank SDPs via manifold optimization

- Degenerate: $\geq$ 2nd relaxation $\rightsquigarrow$ $m \gg n$ **Challenging!**
- Low-rank: $\operatorname{rank} \mathbf{M}^{\mathrm{opt}} \ll n \rightsquigarrow \mathbf{M} = YY^{\intercal},\ Y \in \mathbb{R}^{n\times p}$ **Burer-Monteiro**

  ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p}\}$
- Unital diagonal: $\operatorname{diag}(\mathbf{M}) = \mathbf{1}$

  ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p} \mid \|Y(k,:)\| = 1, k = 1, \ldots, n\}$
- Unital trace: $\operatorname{tr}(\mathbf{M}) = 1$

  ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p} \mid \|Y\|_F = 1\}$

Solving low-rank SDPs via manifold optimization

- Degenerate: $\geq$ 2nd relaxation $\rightsquigarrow$ $m \gg n$ **Challenging!**
- Low-rank: $\operatorname{rank} \mathbf{M}^{\text{opt}} \ll n \rightsquigarrow \mathbf{M} = YY^\intercal,\ Y \in \mathbb{R}^{n\times p}$ **Burer-Monteiro**
  - ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p}\}$
- Unital diagonal: $\operatorname{diag}(\mathbf{M}) = \mathbf{1}$
  - ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p} \mid \|Y(k,:)\| = 1, k = 1, \ldots, n\}$
- Unital trace: $\operatorname{tr}(\mathbf{M}) = 1$
  - ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p} \mid \|Y\|_F = 1\}$

The augmented Lagrangian framework

$$\begin{cases} \inf_{X \succeq 0} & \langle C, X \rangle \\ \text{s.t.} & \mathcal{A}(X) = b, \; \mathcal{B}(X) = d \end{cases}$$

handle with ALM ($\mathcal{A}(X) = b$)

$\mathcal{B}(X) = d$ $\rightsquigarrow$ define a manifold $\mathcal{M}$

- The augmented Lagrangian function:

$$L_\sigma(X, y) = \langle C, X \rangle - y^\intercal(\mathcal{A}(X) - b) + \frac{\sigma}{2}\|\mathcal{A}(X) - b\|^2$$

- Need to solve the subproblem at the k-th step:

$$\min_{X \in \mathcal{M}} L_{\sigma^k}(X, y^k)$$

The augmented Lagrangian framework

$$\begin{cases} \inf_{X \succeq 0} & \langle C, X \rangle \\ \text{s.t.} & \mathcal{A}(X) = b, \ \mathcal{B}(X) = d \end{cases}$$

$\mathcal{A}(X) = b$ → **handle with ALM**

$\mathcal{B}(X) = d$ ⇝ define a manifold $\mathcal{M}$

- The augmented Lagrangian function:

$$L_\sigma(X, y) = \langle C, X \rangle - y^\intercal(\mathcal{A}(X) - b) + \frac{\sigma}{2}\|\mathcal{A}(X) - b\|^2$$

- Need to solve the subproblem at the k-th step:

$$\min_{X \in \mathcal{M}} L_{\sigma^k}(X, y^k)$$

The augmented Lagrangian framework

$$\begin{cases} \inf_{X \succeq 0} & \langle C, X \rangle \\ \text{s.t.} & \mathcal{A}(X) = b, \ \mathcal{B}(X) = d \end{cases}$$

$\mathcal{A}(X) = b$ → **handle with ALM**

$\mathcal{B}(X) = d$ ⇝ define a manifold $\mathcal{M}$

- The augmented Lagrangian function:

$$L_\sigma(X, y) = \langle C, X \rangle - y^\intercal(\mathcal{A}(X) - b) + \frac{\sigma}{2}\|\mathcal{A}(X) - b\|^2$$

- Need to solve the subproblem at the k-th step:

$$\min_{X \in \mathcal{M}} L_{\sigma^k}(X, y^k)$$

Solve the subproblem by the Riemannian trust region method

Let $X = YY^\intercal$. Solve the subproblem on the manifold $\mathcal{N}$ by the Riemannian trust region method:

$$\min_{Y \in \mathcal{N}} \langle C, YY^\intercal \rangle - (y^k)^\intercal(\mathcal{A}(YY^\intercal) - b) + \frac{\sigma^k}{2}\|\mathcal{A}(YY^\intercal) - b\|^2 \rightsquigarrow \text{nonconvex!}$$

Good news

We can efficiently escape from saddle points and arrive at an optimal solution of the original SDP.

Solve the subproblem by the Riemannian trust region method

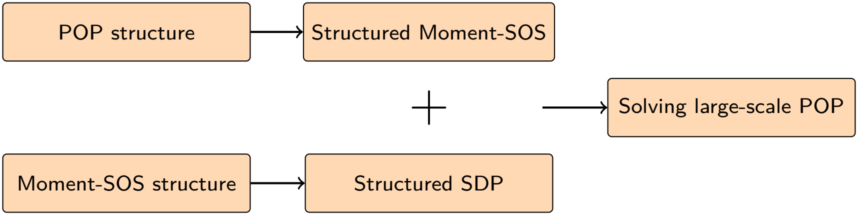
Let $X = YY^\intercal$. Solve the subproblem on the manifold $\mathcal{N}$ by the Riemannian trust region method:

$$\min_{Y \in \mathcal{N}} \langle C, YY^\intercal \rangle - (y^k)^\intercal (\mathcal{A}(YY^\intercal) - b) + \frac{\sigma^k}{2} \|\mathcal{A}(YY^\intercal) - b\|^2 \rightsquigarrow \text{nonconvex!}$$

Good news

We can efficiently escape from saddle points and arrive at an optimal solution of the original SDP.

Solving large-scale polynomial optimization



Software

- TSSOS: based on JuMP, user-friendly, support

commutative/complex/noncommutative polynomial optimization

https://github.com/wangjie212/TSSOS

- ManiSDP: efficiently solve low-rank SDPs via manifold optimization

https://github.com/wangjie212/ManiSDP

Binary quadratic programs

Table: Random binary quadratic programs $\min_{\mathbf{x}\in\{-1,1\}^n} \mathbf{x}^\intercal Q\mathbf{x}$, $r=2$[1]

n	$n_{\rm sdp}$	$m_{\rm sdp}$	MOSEK 10.0		SDPNAL+		STRIDE		ManiSDP	
			$\eta_{\max}$	time	$\eta_{\max}$	time	$\eta_{\max}$	time	$\eta_{\max}$	time
10	56	1,256	4.4e-11	0.71	1.9e-09	0.65	4.7e-13	0.79	3.2e-15	**0.14**
20	211	16,361	2.7e-11	49.0	3.0e-09	28.8	7.4e-13	6.12	1.2e-14	**0.53**
30	466	77,316	-	-	1.7e-04	187	1.2e-12	65.4	2.4e-14	**3.25**
40	821	236,121	-	-	2.1e-08	813	4.4e-13	249	4.1e-14	**10.5**
50	1,276	564,776	-	-	1.6e-07	3058	7.8e-09	826	6.4e-14	**31.1**
60	1,831	1,155,281	-	-	$*$	$*$	1.3e-12	2118	7.9e-14	**94.3**
120	7,261	17,869,161	-	-	-	-	-	-	3.5e-13	**30801**

[1] -: out of memory, $*$: $>$10000s

The robust rotation search problem

- q: unit quaternion parametrization of a 3D rotation
- $(z_i \in \mathbb{R}^3, w_i \in \mathbb{R}^3)_{i=1}^N$: N pairs of 3D points
- $\tilde{z} := [z^\intercal, 0]^\intercal \in \mathbb{R}^4$
- $\tilde{w} := [w^\intercal, 0]^\intercal \in \mathbb{R}^4$
- β_i: threshold determining the maximum inlier residual

$$\min_{\|q\|=1} \sum_{i=1}^{N} \min \left\{ \frac{\|\tilde{z}_i - q \circ \tilde{w}_i \circ q^{-1}\|^2}{\beta_i^2}, 1 \right\}$$

The robust rotation search problem

Table: Results for the robust rotation search problem, $r = 2$

N	MOSEK 10.0		SDPLR 1.03		SDPNAL+		STRIDE		ManiSDP	
	$\eta_{\max}$	time	$\eta_{\max}$	time	$\eta_{\max}$	time	$\eta_{\max}$	time	$\eta_{\max}$	time
50	4.7e-10	16.4	9.8e-03	12.5	1.1e-02	106	2.8e-09	18.3	6.6e-09	**3.02**
100	2.0e-11	622	3.6e-04	106	7.1e-02	642	3.1e-09	73.0	1.0e-09	**22.9**
150	-	-	2.0e-03	291	8.0e-02	1691	4.3e-11	249	1.6e-09	**33.5**
200	-	-	3.1e-02	459	8.3e-02	2799	1.4e-09	254	6.3e-10	**65.3**
300	-	-	1.1e-03	1264	5.2e-02	3528	4.1e-10	1176	1.1e-09	**188**
500	-	-	*	*	*	*	7.1e-09	5627	5.2e-10	**601**

The AC-OPF problem

$$
\begin{cases}
\inf\limits_{V_i, S_k^g \in \mathbb{C}} & \sum\limits_{k \in G} \left(\mathbf{c}_{2k} \Re(S_k^g)^2 + \mathbf{c}_{1k} \Re(S_k^g) + \mathbf{c}_{0k} \right) \\
\text{s.t.} & \angle V_r = 0, \\
& \mathbf{S}_k^{gl} \le S_k^g \le \mathbf{S}_k^{gu}, \quad \forall k \in G, \\
& \boldsymbol{v}_i^l \le |V_i| \le \boldsymbol{v}_i^u, \quad \forall i \in N, \\
& \sum_{k \in G_i} S_k^g - \mathbf{S}_i^d - \mathbf{Y}_i^s |V_i|^2 = \sum_{(i,j) \in E_i \cup E_i^R} S_{ij}, \quad \forall i \in N, \\
& S_{ij} = (\overline{\mathbf{Y}}_{ij} - \mathbf{i}\frac{\mathbf{b}_{ij}^c}{2}) \frac{|V_i|^2}{|\mathbf{T}_{ij}|^2} - \overline{\mathbf{Y}}_{ij} \frac{V_i \overline{V}_j}{\mathbf{T}_{ij}}, \quad \forall (i,j) \in E, \\
& S_{ji} = (\overline{\mathbf{Y}}_{ij} - \mathbf{i}\frac{\mathbf{b}_{ij}^c}{2}) |V_j|^2 - \overline{\mathbf{Y}}_{ij} \frac{\overline{V}_i V_j}{\overline{\mathbf{T}}_{ij}}, \quad \forall (i,j) \in E, \\
& |S_{ij}| \le \mathbf{s}_{ij}^u, \quad \forall (i,j) \in E \cup E^R, \\
& \boldsymbol{\theta}_{ij}^{\Delta l} \le \angle (V_i \overline{V}_j) \le \boldsymbol{\theta}_{ij}^{\Delta u}, \quad \forall (i,j) \in E.
\end{cases}
$$

The AC-OPF problem

n	m	CS ($r = 2$)				CS+TS ($r = 2$)			
		n_{sdp}	opt	time	gap	n_{sdp}	opt	time	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%

Nonlinear Bell inequality

- $\lambda(A_1B_2 + A_2B_1)^2 + \lambda(A_1B_1 - A_2B_2)^2 \le 4$

$$\begin{cases} \sup_{x_i,y_j} & (\varsigma(x_1y_2) + \varsigma(x_2y_1))^2 + (\varsigma(x_1y_1) - \varsigma(x_2y_2))^2 \\ \text{s.t.} & x_i^2 = 1, y_j^2 = 1, [x_i, y_j] = 0 \text{ for } i, j = 1, 2. \end{cases}$$

- For classical models: 4
- For quantum commuting model: 4 ($r = 3$)

Nonlinear Bell inequality

- $\lambda(A_1B_2 + A_2B_1)^2 + \lambda(A_1B_1 - A_2B_2)^2 \leq 4$

$$\begin{cases} \sup\limits_{x_i,y_j} & (\varsigma(x_1y_2) + \varsigma(x_2y_1))^2 + (\varsigma(x_1y_1) - \varsigma(x_2y_2))^2 \\ \text{s.t.} & x_i^2 = 1, y_j^2 = 1, [x_i, y_j] = 0 \text{ for } i, j = 1, 2. \end{cases}$$

- For classical models: 4
- For quantum commuting model: 4 ($r = 3$)

Nonlinear Bell inequality

- $\lambda(A_2 + B_1 + B_2 - A_1B_1 + A_2B_1 + A_1B_2 + A_2B_2) - \lambda(A_1)\lambda(B_1) - \lambda(A_2)\lambda(B_1) - \lambda(A_2)\lambda(B_2) - \lambda(A_1)^2 - \lambda(B_2)^2$

$$\begin{cases} \sup\limits_{x_i,y_j} & \varsigma(x_2) + \varsigma(y_1) + \varsigma(y_2) - \varsigma(x_1y_1) + \varsigma(x_2y_1) + \varsigma(x_1y_2) + \varsigma(x_2y_2) \\ & -\varsigma(x_1)\varsigma(y_1) - \varsigma(x_2)\varsigma(y_1) - \varsigma(x_2)\varsigma(y_2) - \varsigma(x_1)^2 - \varsigma(y_2)^2 \\ \text{s.t.} & x_i^2 = 1, y_j^2 = 1, [x_i, y_j] = 0 \text{ for } i, j = 1, 2. \end{cases}$$

- For classical models: 3.375
- For quantum commuting model: 3.5114 ($r = 2$)

Nonlinear Bell inequality

- $\lambda(A_2 + B_1 + B_2 - A_1B_1 + A_2B_1 + A_1B_2 + A_2B_2) - \lambda(A_1)\lambda(B_1) - \lambda(A_2)\lambda(B_1) - \lambda(A_2)\lambda(B_2) - \lambda(A_1)^2 - \lambda(B_2)^2$

$$\begin{cases} \sup\limits_{x_i,y_j} & \varsigma(x_2)+\varsigma(y_1)+\varsigma(y_2)-\varsigma(x_1y_1)+\varsigma(x_2y_1)+\varsigma(x_1y_2)+\varsigma(x_2y_2) \\ & -\varsigma(x_1)\varsigma(y_1)-\varsigma(x_2)\varsigma(y_1)-\varsigma(x_2)\varsigma(y_2)-\varsigma(x_1)^2-\varsigma(y_2)^2 \\ \text{s.t.} & x_i^2=1, y_j^2=1, [x_i,y_j]=0 \text{ for } i,j=1,2. \end{cases}$$

- For classical models: 3.375
- For quantum commuting model: 3.5114 ($r = 2$)

Ground state energy of quantum many-body systems

The Heisenberg chain is defined by the Hamiltonian:

$$H = \sum_{i=1}^{N} \sum_{a\in\{x,y,z\}} \sigma_i^a \sigma_{i+1}^a.$$

The ground state energy of the Heisenberg chain equals the optimum of the NCPOP:

$$\begin{cases} \min\limits_{\{|\psi\rangle,\sigma_i^a\}} & \langle\psi|H|\psi\rangle \\ \text{s.t.} & (\sigma_i^a)^2 = 1, \quad i = 1,\ldots,N, a \in \{x,y,z\}, \\ & \sigma_i^x\sigma_i^y = \mathbf{i}\sigma_i^z, \sigma_i^y\sigma_i^z = \mathbf{i}\sigma_i^x, \sigma_i^z\sigma_i^x = \mathbf{i}\sigma_i^y, \quad i = 1,\ldots,N, \\ & \sigma_i^a\sigma_j^b = \sigma_j^b\sigma_i^a, \quad 1 \le i \ne j \le N, a, b \in \{x,y,z\}. \end{cases}$$

Ground state energy of many-body systems

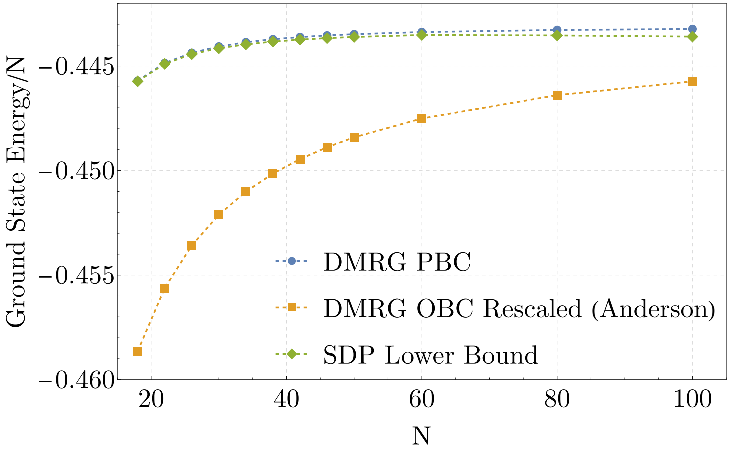
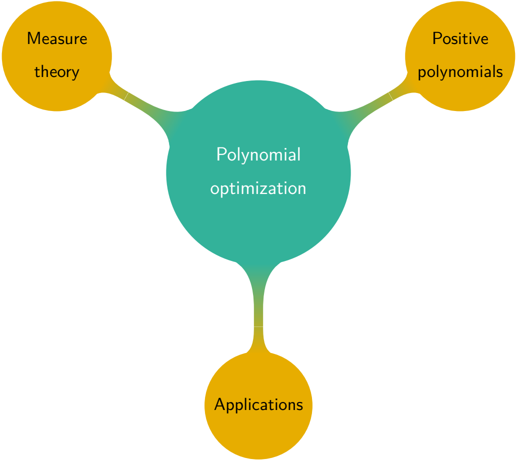


Figure: Ground state energy of the Heisenberg chain

Summary


Measure
theory
Positive
polynomials
Polynomial
optimization
Applications

Conclusions

- Polynomial optimization provides a unified scheme for global optimization of various non-convex problems.
- The scalability of the moment-SOS hierarchy can be significantly improved by exploiting plenty of algebraic structures.
- There are tons of applications in diverse fields!

Main references

- **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** and Victor Magron, Exploiting Sparsity in Complex Polynomial Optimization, Journal of Optimization Theory and Applications, 2021.
- **Jie Wang**, Victor Magron, Jean B. Lasserre and Ngoc H. A. Mai, *CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization*, ACM Transactions on Mathematical Software, 2022.

Main references

- **Jie Wang** and Victor Magron, Exploiting Term Sparsity in Noncommutative Polynomial Optimization, Computational Optimization and Applications, 2021.
- Igor Klep, Victor Magron, Jurij Volčič and **Jie Wang**, State Polynomials: Positivity, Optimization and Nonlinear Bell Inequalities, arXiv, 2023.
- **Jie Wang** and Liangbing Hu, Solving Low-Rank Semidefinite Programs via Manifold Optimization, arXiv, 2023.
- Feng Guo and **Jie Wang**, A Moment-SOS Hierarchy for Robust Polynomial Matrix Inequality Optimization with SOS-Convexity, arXiv, 2023.

Many applications, including computer vision, computer arithmetic, deep learning, entanglement in quantum information, graph theory and energy networks, can be successfully tackled within the framework of polynomial optimization, an emerging field with growing research efforts in the last two decades. One key advantage of these techniques is their ability to model a wide range of problems using optimization formulations. Polynomial optimization heavily relies on the moment-sums of squares (moment-SOS) approach proposed by Lasserre, which provides certificates for positive polynomials. On the practical side, however, there is "no free lunch" and such optimization methods usually encompass severe scalability issues. Fortunately, for many applications, including the ones formerly mentioned, we can look at the problem in the eyes and exploit the inherent data structure arising from the cost and constraints describing the problem.
This book presents several research efforts to resolve this scientific challenge with important computational implications. It provides the development of alternative optimization schemes that scale well in terms of computational complexity, at least in some identified class of problems. It also features a unified modeling framework to handle a wide range of applications involving both commutative and noncommutative variables, and solves concretely large-scale instances. Readers will find a practical section dedicated to the use of available open-source software libraries.
This interdisciplinary monograph is essential reading for students, researchers and professionals interested in solving optimization problems with polynomial input data.
World Scientific
www.worldscientific.com
Q0382 hc
ISSN 2399-1593
ISBN 978-1-80061-294-5
Vol. 5
Sparse Polynomial Optimization
Magron
Wang
Series on Optimization and its Applications – Vol. 5
Sparse Polynomial Optimization
Theory and Practice
Victor Magron
Jie Wang
World Scientific

Thank You!

https://wangjie212.github.io/jiewang