Polynomial Optimization and Low-Rank SDPs

Jie Wang

Academy of Mathematics and Systems Science, CAS

June 1, 2023

Collaborators

Jean B. Lasserre

Victor Magron

Outline

1. Polynomial optimization and the moment-SOS hierarchy

2. Reducing complexity by exploiting sparsity

3. Efficiently solving low-rank SDPs

Outline

1. Polynomial optimization and the moment-SOS hierarchy
2. Reducing complexity by exploiting sparsity
3. Efficiently solving low-rank SDPs

Outline

1. Polynomial optimization and the moment-SOS hierarchy

2. Reducing complexity by exploiting sparsity

3. Efficiently solving low-rank SDPs

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?

- Powerful modelling ability: QCQP, binary programs, mixed integer nonlinear programs and so on
- 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

Why polynomial optimization?

- Powerful modelling ability: QCQP, binary programs, mixed integer nonlinear programs and so on
- 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

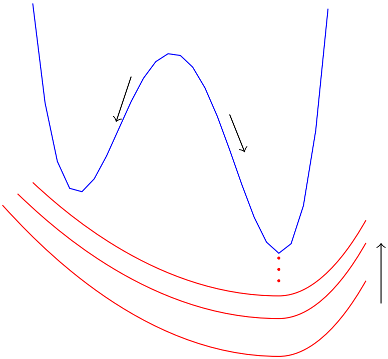
Why polynomial optimization?

- Powerful modelling ability: QCQP, binary programs, mixed integer nonlinear programs and so on
- 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

Why polynomial optimization?

- Powerful modelling ability: QCQP, binary programs, mixed integer nonlinear programs and so on
- 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

Polynomial optimization



Example (moment relaxation)

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

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

The hierarchy of moment relaxations

- The hierarchy of moment relaxations (Lasserre 2001):

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

Example (SOS relaxation)

$$\begin{cases}\inf\limits_{\mathbf{x}} & x_1^2+x_1x_2+x_2^2\\ \text{s.t.} & 1-x_1^2\ge 0, 1-x_2^2\ge 0\end{cases} \Longleftrightarrow \begin{cases}\sup\limits_{\lambda} & \lambda\\ \text{s.t.} & x_1^2+x_1x_2+x_2^2-\lambda\ge 0, \forall\mathbf{x}\in\mathbb{R}^2 \text{ s.t. } (1-x_1^2\ge 0, 1-x_2^2\ge 0)\end{cases}$$

$$\overset{\text{relax}}{\Longrightarrow} \textbf{(SOS)} \begin{cases}\sup\limits_{\lambda} & \lambda\\ \text{s.t.} & x_1^2+x_1x_2+x_2^2-\lambda=\sigma_0+\sigma_1(1-x_1^2)+\sigma_2(1-x_2^2),\\ & \sigma_0,\sigma_1,\sigma_2\in\text{SOS}\end{cases}$$

The hierarchy of dual SOS relaxations

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

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

The Moment-SOS/Lasserre's hierarchy

$$\begin{array}{ccccccc}
 & & & f_{\min} & & & \\
 & & \rotatebox{45}{$\leq$} & & \rotatebox{135}{$\leq$} & & \\
 & & \vdots & & \vdots & & \\
 & & \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} & & \\
\text{(Moment relaxation)} & & \theta_r & \text{``}=\text{''} & \theta_r^* & & \text{(SOS relaxation)} \\
 & & \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} & & \\
 & & \vdots & & \vdots & & \\
 & & \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} & & \\
 & & \theta_{r_{\min}} & \text{``}=\text{''} & \theta_{r_{\min}}^* & &
\end{array}$$

Asymptotical convergence and finite convergence

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

➤ $\theta_r \uparrow f_{\min}$ and $\theta_r^* \uparrow f_{\min}$ 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. $N - ||\mathbf{x}||^2 \in \mathcal{Q}(\mathbf{g})$
  - ➤ $\theta_r \uparrow f_{\min}$ and $\theta_r^* \uparrow f_{\min}$ 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. $N - ||\mathbf{x}||^2 \in \mathcal{Q}(\mathbf{g})$
  - ➤ $\theta_r \uparrow f_{\min}$ and $\theta_r^* \uparrow f_{\min}$ 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 $r_{\min} \le r' \le r$, $\operatorname{rank} \mathbf{M}_{r'-r_{\min}}(\mathbf{y}) = \operatorname{rank} \mathbf{M}_{r'}(\mathbf{y})$
    - $\rightsquigarrow$ 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 $r_{\min} \le r' \le r$, $\operatorname{rank} \mathbf{M}_{r'-r_{\min}}(\mathbf{y}) = \operatorname{rank} \mathbf{M}_{r'}(\mathbf{y})$
    - ⇝ Extract $\operatorname{rank} \mathbf{M}_{r'}(\mathbf{y})$ globally optimal solutions
  - $\operatorname{rank} \mathbf{M}_{r_{\min}}(\mathbf{y}) = 1$
    - ⇝ 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 $r_{\min} \le r' \le r$, $\mathrm{rank}\,\mathbf{M}_{r'-r_{\min}}(\mathbf{y}) = \mathrm{rank}\,\mathbf{M}_{r'}(\mathbf{y})$
    - $\rightsquigarrow$ 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

Computational bottleneck of the Moment-SOS hierarchy

- The size of SDP corresponding to the r-th order SOS relaxation:
  1. PSD constraint: $\binom{n+r}{r}$
  2. #equality constraint: $\binom{n+2r}{2r}$
- $r = 2$, $n < 30$ (Mosek)
- Exploiting structure:
  - quotient ring
  - symmetry
  - sparsity

Computational bottleneck of the Moment-SOS hierarchy

- The size of SDP corresponding to the r-th order SOS relaxation:
  1. PSD constraint: $\binom{n+r}{r}$
  2. #equality constraint: $\binom{n+2r}{2r}$
- $r = 2,\ n < 30$ (Mosek)
- Exploiting structure:
  - quotient ring
  - symmetry
  - sparsity

Computational bottleneck of the Moment-SOS hierarchy

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

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

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

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

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

Term sparsity (Wang & Magron & Lasserre, 2021)

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

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

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

Extensions

- Complex polynomial optimization $\rightsquigarrow$ optimal power flow
- Trigonometric polynomial optimization $\rightsquigarrow$ sigal processing
- Noncommutative polynomial optimization $\rightsquigarrow$ Bell inequality, quantum entanglement, quantum polynomial optimization
- Polynomial matrix optimization $\rightsquigarrow$ maximal/minimal eigenvalue
- Computation of joint spectral radius $\rightsquigarrow$ stability of switched linear system
- Polynomial dynamic system $\rightsquigarrow$ maximal controlled invariant set, attraction region, global attractor, reachable set, optimal control

Software

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

commutative/complex/noncommutative polynomial optimization

https://github.com/wangjie212/TSSOS

The AC-OPF problem

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

The AC-OPF problem

n	m	CS ($r = 2$)				CS+TS ($r = 2, s = 1$)			
		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%

Solving low-rank SDPs via manifold optimization

- Degenerate: $\geq$ 2nd order 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 order 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 order 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 order 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: $\mathrm{tr}(\mathbf{M}) = 1$
  - ➤ $\mathcal{N} := \{Y \in \mathbb{R}^{n\times p} \mid \|Y\|_F = 1\}$

The augmented Lagrangian framework

$$\begin{cases}\inf\limits_{X\succeq 0} & \langle C, X\rangle \\ \text{s.t.} & \mathcal{A}(X) = b, \ \mathcal{B}(X) = d \rightsquigarrow \text{define a manifold } \mathcal{M}\end{cases}$$

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

- 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

$$\begin{cases} \inf\limits_{X\succeq 0} & \langle C, X\rangle \\ \text{s.t.} & \mathcal{A}(X) = b,\ \mathcal{B}(X) = d \rightsquigarrow \text{define a manifold } \mathcal{M} \end{cases}$$

handle with ALM

- 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

$$\begin{cases} \inf\limits_{X\succeq 0} & \langle C, X\rangle \\ \text{s.t.} & \overset{\hookrightarrow\ \textbf{handle with ALM}}{\mathcal{A}(X)=b}, \ \mathcal{B}(X)=d \ \rightsquigarrow \text{define a manifold } \mathcal{M} \end{cases}$$

- 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 obtain an optimal solution of the 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 obtain an optimal solution of the SDP.

Numerical experiments

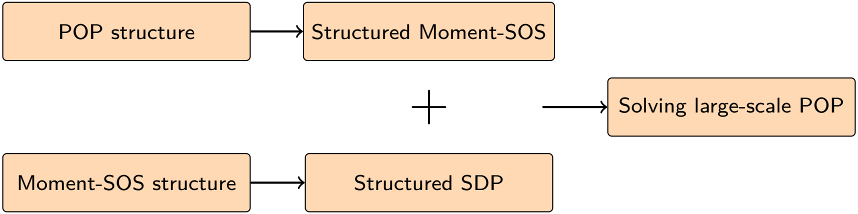
Table: Binary quadratic programs $\min_{\mathbf{x}\in\{-1,1\}^d} \mathbf{x}Q\mathbf{x}^\intercal$, $r = 2$[1]

d	n	m	`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**

---

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

Solving large-scale polynomial optimization


POP structure

Structured Moment-SOS

Moment-SOS structure

Structured SDP

Solving large-scale POP

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.

Thank You!

https://wangjie212.github.io/jiewang