Exploiting Sparsity in Large-Scale Polynomial Optimization

Jie Wang (AMSS-CAS)
Joint work with Victor Magron and Jean B. Lasserre

SIAM Conference on Optimization

21/7/2021

Content

1. Polynomial optimization and the moment-SOS hierarchy
2. Exploiting sparsity in the moment-SOS hierarchy
3. Numerical experiments

Polynomial optimization problem

We consider the polynomial optimization problem (POP):

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

where $f, g_j\,(, h_i) \in \mathbb{R}[\mathbf{x}] := \mathbb{R}[x_1, \ldots, x_n]$.

In general, the problem (Q) is non-convex, NP-hard.

Diverse applications

- Combinatorial optimization (e.g. the Max-Cut problem)
- Signal processing
- Tensor decomposition
- Optimal power flow
- Computer vision
- Neural networks
- Quantum information
- ......

What do we care about?

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

What do we care about?

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

The moment-SOS hierarchy (also known as Lasserre's hierarchy) is a well-established tool to handle POPs and is able to answer all these questions.

Moment matrix and localizing matrix

For $r \in \mathbb{N}$, let $\mathbb{N}^n_r := \{\boldsymbol{\beta} = (\beta_i) \in \mathbb{N}^n \mid \sum_{i=1}^n \beta_i \le r\}$ arranged w.r.t. the lexicographic order.

▷ The moment matrix $M_r(\mathbf{y})$ of order r is defined by

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

▷ Given $g = \sum_{\boldsymbol{\alpha}} g_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}} \in \mathbb{R}[\mathbf{x}]$, the localizing matrix $M_r(g\mathbf{y})$ of order r is defined by

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

The moment relaxation

By truncating the order of moments, for each r (called the relaxation order), we obtain the moment relaxation of order r:

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

Here, $d_j = \lceil \deg(g_j)/2 \rceil$.

This is a semidefinite programming (SDP) problem, effectively solved by interior-point solvers (e.g. MOSEK) or first-order solvers (e.g. COSMO).

The dual SOS relaxation

The dual SDP of the moment relaxation of order r is the following SOS relaxation of order r:

$$
(\mathrm{Q}_r)^*: \quad \begin{array}{rll} \theta_r^* := & \sup & \lambda \\ & \text{s.t.} & f - \lambda = \sigma_0 + \sum_{j=1}^m \sigma_j g_j, \\ & & \sigma_0, \sigma_1, \ldots, \sigma_m \text{ are SOS polynomials}, \\ & & \deg(\sigma_0) \le 2r, \deg(\sigma_j g_j) \le 2r, j = 1, \ldots, m. \end{array}
$$

$$
\begin{array}{ccccc}
 & & f^* & & \\
 & \rotatebox{-45}{$\geq$} & & \rotatebox{45}{$\leq$} & \\
 & \vdots & & \vdots & \\
 & \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} & \\
\text{(the moment relaxation)} & \theta_r & \text{“}=\text{”} & \theta_r^* & \text{(the SOS relaxation)} \\
 & \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} & \\
 & \vdots & & \vdots & \\
 & \rotatebox{90}{$\leq$} & & \rotatebox{90}{$\leq$} & \\
 & \theta_{\underline{r}} & \text{“}=\text{”} & \theta_{\underline{r}}^* &
\end{array}
$$

$$\underline{r} := \max\{\deg(f)/2, d_1, \ldots, d_m\}$$

Asymptotical convergence and finite convergence

Under Archimedean's condition: there exists $N > 0$ s.t. $N - ||\mathbf{x}||^2 \in \mathcal{Q}_{\mathbf{g}}$, we have

- $\theta_r \uparrow f^*$ and $\theta_r^* \uparrow f^*$ as $r \to \infty$ (Lassere, 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.

In practice for most POPs, the moment-SOS hierarchy retrieves f^* in a few steps.

Asymptotical convergence and finite convergence

Under Archimedean's condition: there exists $N > 0$ s.t. $N - ||\mathbf{x}||^2 \in \mathcal{Q}_{\mathbf{g}}$, we have

- $\theta_r \uparrow f^*$ and $\theta_r^* \uparrow f^*$ as $r \to \infty$ (Lassere, 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.

In practice for most POPs, the moment-SOS hierarchy retrieves f^* in a few steps.

Key Message: The moment-SOS hierarchy allows us to approximate/retrieve the global optimum/optimizers via solving a sequence of SDPs with increasing sizes.

Scalability issue

The size of SDP (considering the SOS problem) at relaxation order r:

- maximal size of SDP matrices: $\binom{n+r}{r}$
- number of equality constraints: $\binom{n+2r}{2r}$

In view of the current state of SDP solvers (e.g. MOSEK), solvable problems are limited to $n \leq 30$ when $r = 2$ on a standard laptop.

Scalability issue

The size of SDP (considering the SOS problem) at relaxation order r:

- maximal size of SDP matrices: $\binom{n+r}{r}$
- number of equality constraints: $\binom{n+2r}{2r}$

In view of the current state of SDP solvers (e.g. MOSEK), solvable problems are limited to $n \leq 30$ when $r = 2$ on a standard laptop.

Exploiting structure:

- quotient ring
- symmetry
- sparsity

Correlative sparsity (Waki et al., 2006)

The basic idea is to partition the variables into cliques according to the correlations between variables.

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

Correlative sparsity (Waki et al., 2006)

The basic idea is to partition the variables into cliques according to the correlations between variables.

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

We then construct moment/localizing matrices with respect to the variables involved in each maximal clique of the csp graph:

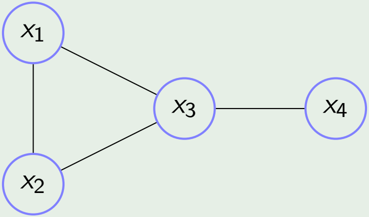
$$I_k \longmapsto M_r(\mathbf{y}, I_k), M_{r-d_j}(g_j\mathbf{y}, I_k)$$

Correlative sparsity

Example

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

Figure: The csp graph for f and $\{g_1, g_2\}$



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

The correlative sparsity adapted moment-SOS hierarchy

- If the csp graph is chordal (otherwise we need a chordal extension), then the correlative sparsity adapted moment-SOS hierarchy 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

In contrast with correlative sparsity concerning variables, term sparsity treats sparsity at the term/monomial level.

Term sparsity

In contrast with correlative sparsity concerning variables, term sparsity treats sparsity at the term/monomial level.

$V_r(\mathbf{x}) := \{1, x_1, \ldots, x_n, x_1^r, \ldots, x_n^r\}$ the monomial basis of degree $\le r$.

Term sparsity pattern (tsp) graph $G^{\mathrm{tsp}}(V, E)$ (with relaxation order r):
$V := V_r(\mathbf{x})$
$\{\mathbf{x}^{\boldsymbol{\alpha}}, \mathbf{x}^{\boldsymbol{\beta}}\} \in E \iff \mathbf{x}^{\boldsymbol{\alpha}} \cdot \mathbf{x}^{\boldsymbol{\beta}} = \mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \in \mathrm{supp}(f) \cup \bigcup_{j=1}^{m} \mathrm{supp}(g_j) \cup V_r(\mathbf{x})^2$

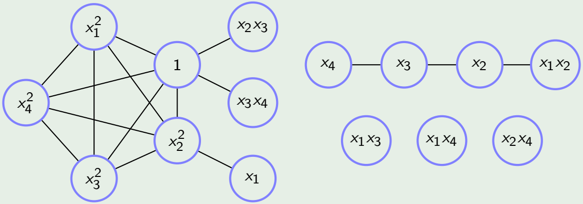
(For $f = \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}} \in \mathbb{R}[\mathbf{x}]$, $\mathrm{supp}(f) := \{\mathbf{x}^{\boldsymbol{\alpha}} \mid f_{\boldsymbol{\alpha}} \neq 0\}$)

Term sparsity

Example

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

Figure: The tsp graph for f and $\{g_1, g_2\}$ with $r = 2$



Term sparsity

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

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

Term sparsity

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

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

In the moment relaxation,

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

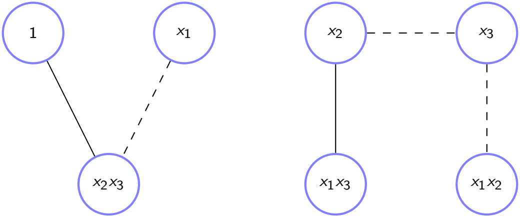
Similarly for the localizing matrices $M_{r-d_j}(\mathbf{y}), j = 1, \ldots, m$.

Extending to an iterative procedure

By iteratively performing support extension and chordal extension:

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

Figure: Support extension ($\mathbf{x}^{\boldsymbol{\beta}'}\mathbf{x}^{\gamma'} = \mathbf{x}^{\boldsymbol{\beta}}\mathbf{x}^{\gamma}$ and $\{\mathbf{x}^{\boldsymbol{\beta}}, \mathbf{x}^{\gamma}\} \in E \Rightarrow \{\mathbf{x}^{\boldsymbol{\beta}'}, \mathbf{x}^{\gamma'}\} \in E$)



The term sparsity adapted moment-SOHS hierarchy

Let $C_{j,1}^{(s)}, \ldots, C_{j,t_{j,s}}^{(s)}$ be the maximal cliques of $G_j^{(s)}$. For each $s \geq 1$, let us consider

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

We call $\{(\mathrm{Q}_r^s)\}_{r,s}$ the TSSOS hierarchy and s the sparse order.

A two-level hierarchy of lower bounds

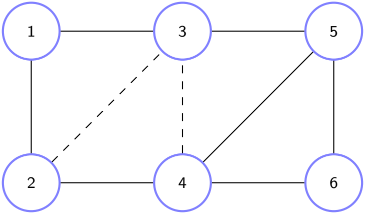
Consequently, we obtain a two-level hierarchy of lower bounds for f^*:
$(\underline{r} := \max\{\deg(f)/2, d_1, \ldots, d_m\})$

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

Different choices of chordal extensions

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

Figure: smallest versus maximal chordal extension



The TSSOS hierarchy

Regarding the TSSOS hierarchy, we have

- For QCQP, $\theta_1^{(1)} = \theta_{\mathrm{shor}}$;
- Fixing a sparse order s, the sequence $(\theta_r^{(s)})_{r\geq\underline{r}}$ is monotonically nondecreasing;
- Fixing a relaxation order r, the sequence $(\theta_r^{(s)})_{s\geq 1}$ is monotonically nondecreasing and converges to θ_r in finitely many steps if the maximal chordal extension is used for the chordal extension operation.

Combining correlative sparsity with term sparsity

The combination of correlative sparsity with term sparsity splits into two steps:

1. Partitioning 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.

In doing so, we again obtain a two-level hierarchy of lower bounds for f^*, which is called the CS-TSSOS hierarchy.

Software

The sparsity-adapted hierarchies have been implemented in the Julia package TSSOS (freely available on GitHub):

https://github.com/wangjie212/TSSOS

Randomly generated polynomials of the SOS form

TSSOS, GloptiPoly, Yalmip: MOSEK   SparsePOP: SDPT3

Table: Running time (in seconds) comparison with GloptiPoly, Yalmip and SparsePOP for minimizing randomly generated sparse polynomials of the SOS form with the same optimum; the symbol "-" indicates out of memory

n	$2d$	TSSOS	GloptiPoly	Yalmip	SparsePOP
8	8	0.24	306	10	24
8	8	0.34	348	13	130
8	8	0.36	326	19	175
8	10	0.58	-	92	323
8	10	0.53	-	72	1526
8	10	0.38	-	22	134
9	10	0.50	-	44	324
9	10	0.72	-	143	-
9	10	0.79	-	109	284
10	12	2.2	-	474	-
10	12	1.6	-	147	318
10	12	1.8	-	350	404
10	16	15	-	-	-
10	16	14	-	-	-
10	16	12	-	-	-
12	12	8.4	-	-	-
12	12	5.7	-	-	-
12	12	7.4	-	-	-

Randomly generated polynomials with simplex Newton polytopes

Table: Running time (in seconds) comparison with GloptiPoly, Yalmip and SparsePOP for minimizing randomly generated sparse polynomials with simplex Newton polytopes with the same optimum; the symbol "-" indicates out of memory

n	$2d$	TSSOS	GloptiPoly	Yalmip	SparsePOP
8	8	0.36	346	31	271
8	8	0.51	447	24	496
8	8	0.31	257	6.0	178
9	8	1.0	-	-	-
9	8	0.63	-	363	611
9	8	0.76	-	141	578
9	10	6.6	-	322	-
9	10	5.0	-	233	-
9	10	4.9	-	249	-
10	8	1.2	-	-	-
10	8	8.0	-	536	-
10	8	1.0	-	-	-
11	8	1.7	-	655	398
11	8	1.8	-	-	221
11	8	1.9	-	340	293
12	8	10	-	-	-
12	8	7.4	-	-	-
12	8	2.9	-	-	-

The AC optimal power flow problem

Table: The results for AC-OPF instances; mb: the maximal size of blocks, gap: the optimality gap with respect to a local optimal solution, -: out of memory

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%
114	315	66	1.3442e5	5.59	0.39%	31	1.3396e5	2.01	0.73%
114	315	120	7.6943e4	94.9	0.00%	39	7.6942e4	14.8	0.00%
72	297	45	4.9927e3	4.43	0.07%	22	4.9920e3	2.69	0.08%
344	971	153	4.2246e5	758	0.06%	44	4.2072e5	96.0	0.48%
344	971	153	2.2775e5	504	0.00%	44	2.2766e5	71.5	0.04%
344	1325	253	—	—	—	31	2.4180e5	82.7	0.11%
344	1325	253	—	—	—	73	1.0470e5	169	0.50%
348	1809	253	—	—	—	34	1.0802e5	278	0.05%
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	231	4.2413e4	3114	0.85%	39	4.2408e4	46.6	0.86%
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%

Conclusion and outlook

- The concept of term sparsity patterns opens a new window to exploit sparsity at the term level for polynomial optimization;
- When appropriate sparsity patterns are accessible, we can significantly improve the scalability of the moment-SOS hierarchy;
- Extensions to other situations also relying on the moment-SOS hierarchy, e.g., complex polynomial optimization, noncommutative polynomial optimization;
- Fruitful potential applications: optimal power flow, computer vision, neural networks, quantum information, control and so on.

Main references

- **Jie Wang**, Victor Magron and Jean B. Lasserre, *TSSOS: A Moment-SOS hierarchy that exploits term sparsity*, SIAM Optimization, 2021.
- **Jie Wang**, Victor Magron and Jean B. Lasserre, *Chordal-TSSOS: a moment-SOS hierarchy that exploits term sparsity with chordal extension*, SIAM 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.
- **Jie Wang** and Victor Magron, Exploiting Term Sparsity in Noncommutative Polynomial Optimization, Computational Optimization and Applications, 2021.
- **Jie Wang**, Martina Maggio and Victor Magron, *SparseJSR: A Fast Algorithm to Compute Joint Spectral Radius via Sparse SOS Decompositions*, American Control Conference, 2021.

- `TSSOS`: https://github.com/wangjie212/TSSOS

Thanks for your attention!