Exploiting Term Sparsity in Large-Scale Polynomial Optimization

Jie Wang
Joint work with Victor Magron and Jean B. Lasserre

LAAS-CNRS

11/9/2020

Polynomial optimization problem

Let us consider the polynomial optimization problem:

$$(\mathrm{Q}): \quad \begin{aligned} \theta^* := \; & \inf \quad f \\ & \text{s.t.} \quad g_j \geq 0, \quad j = 1, \ldots, m, \end{aligned}$$

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

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

Moment-SOS hierarchies

The moment-SOS hierarchy (also known as Lasserre's hierarchy) defines a series of SDP relaxations to approximate θ^* from below:

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

with the dual SDP

$$(\mathrm{Q}_d)^*: \quad \begin{array}{ll} \sup & \lambda \\ \text{s.t.} & f - \lambda = s_0 + \sum_{j=1}^m s_j g_j, \\ & s_j \in \Sigma_{2(d-d_j)}, \quad j = 0, \ldots, m. \end{array}$$

Here, $d_0 = 0, d_j = \lceil \deg(g_j)/2 \rceil$.

Asymptotical convergence and finite Convergence

Under Archimedean's condition: there exists an $N > 0$ such that
$N - ||\mathbf{x}||^2 \in \mathcal{Q}(g_1, \ldots, g_m) := \{s_0 + \sum_{j=1}^{m} s_j g_j \mid s_j \in \Sigma, j = 0, \ldots, m\}$,

- $\theta_d \uparrow \theta^*$ as $d \to \infty$;
- Finite convergence happens generically (Nie, 2014).

Asymptotical convergence and finite Convergence

Under Archimedean's condition: there exists an $N > 0$ such that $N - ||\mathbf{x}||^2 \in \mathcal{Q}(g_1, \ldots, g_m) := \{s_0 + \sum_{j=1}^m s_j g_j \mid s_j \in \Sigma, j = 0, \ldots, m\}$,

- $\theta_d \uparrow \theta^*$ as $d \to \infty$;
- Finite convergence happens generically (Nie, 2014).

So the moment-SOS hierarchy enable us to approximate/retrieve θ^* via solving a sequence of SDPs with increasing sizes.

Scalability

The size of SDP (considering $(\mathrm{Q}_d)^*$) at relaxation order d:

- PSD matrix: $\binom{n+d}{d}$
- #equality constraint: $\binom{n+2d}{2d}$

Given the current state of SDP solvers (e.g. Mosek), problems are limited to $n < 30$ when $d = 2$.

Scalability

The size of SDP (considering $(\mathrm{Q}_d)^*$) at relaxation order d:

- PSD matrix: $\binom{n+d}{d}$
- #equality constraint: $\binom{n+2d}{2d}$

Given the current state of SDP solvers (e.g. Mosek), problems are limited to $n < 30$ when $d = 2$.

Exploiting structure: quotient ring, symmetry, sparsity

Correlative sparsity (Waki et al., 2006)

The basic idea is to partition the variables into groups according to the correlation between variables.

Correlative sparsity pattern (csp) graph $G^{\mathrm{csp}}(V, E)$:
$V := \{x_1, \ldots, x_n\}$
$\{x_i, x_j\} \in E \iff x_i, x_j$ appear in the same term of f or appear in the same constraint g_j

Correlative sparsity (Waki et al., 2006)

The basic idea is to partition the variables into groups according to the correlation between variables.

Correlative sparsity pattern (csp) graph $G^{\mathrm{csp}}(V, E)$:
$V := \{x_1, \ldots, x_n\}$
$\{x_i, x_j\} \in E \iff x_i, x_j$ appear in the same term of f or appear in the same constraint g_j

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

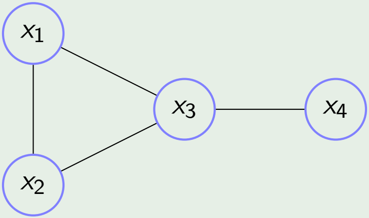
$$I_k \longmapsto M_d(\mathbf{y}, I_k), M_{d-d_j}(g_j\mathbf{y}, I_k)$$

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 $\{g_1, g_2\}$



Two maximal cliques: $\{x_1, x_2, x_3\}$ and $\{x_3, x_4\}$

Term sparsity

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

Term sparsity

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

$V_d(\mathbf{x}) := \{1, x_1, \ldots, x_n, x_1^d, \ldots, x_n^d\}$ the monomial basis of degree $\leq d$.

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

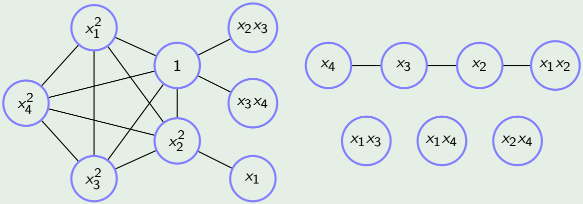
(For $f = \sum_{\boldsymbol{\alpha}} f_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}} \in \mathbb{R}[\mathbf{x}]$, $\mathrm{supp}(f) := \{\mathbf{x}^{\boldsymbol{\alpha}} \mid f_{\boldsymbol{\alpha}} \neq 0\}$)

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 $d = 2$



Term sparsity

Suppose the tsp graph G^{tsp} has connected components: $\mathscr{B}_1, \ldots, \mathscr{B}_t$. So

$$V_d(\mathbf{x}) = \bigsqcup_{i=1}^{t} \mathscr{B}_i.$$

For each $\mathscr{B}_i$, we construct a block of the moment matrix: $M_{\mathscr{B}_i}(\mathbf{y})$.

Term sparsity

Suppose the tsp graph G^{tsp} has connected components: $\mathscr{B}_1, \ldots, \mathscr{B}_t$. So

$$V_d(\mathbf{x}) = \bigsqcup_{i=1}^{t} \mathscr{B}_i.$$

For each $\mathscr{B}_i$, we construct a block of the moment matrix: $M_{\mathscr{B}_i}(\mathbf{y})$.

In such a way, we replace one big matrix $M_d(\mathbf{y})$ by a series of smaller matrices $M_{\mathscr{B}_i}(\mathbf{y}), i = 1, \ldots, t$ in the moment relaxation.

Remark: The same thing can be also done for the localizing matrices $M_{d-d_j}(\mathbf{y}), j = 1, \ldots, m$.

Extending to an iterative procedure

For simplicity, we consider the unconstrained case. For a graph $G(V, E)$ with nodes $V_d(\mathbf{x})$ ($d = \deg(f)/2$), define

$$\mathrm{supp}(G) := \{\mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \mid \{\mathbf{x}^{\boldsymbol{\alpha}}, \mathbf{x}^{\boldsymbol{\beta}}\} \in E\}.$$

Let $G^{(0)} = G^{\mathrm{tsp}}$. We iteratively define a sequence of graphs $(G^{(k)})_{k\geq 1}$ via two successive steps:

1. Support-extension: let $F^{(k)}$ be the graph with nodes $V_d(\mathbf{x})$ and edges

$$E(F^{(k)}) := \{\{\mathbf{x}^{\boldsymbol{\alpha}}, \mathbf{x}^{\boldsymbol{\beta}}\} \mid \mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \in \mathrm{supp}(G^{(k-1)}) \cup (2V_d(\mathbf{x}))\}$$

2. Block-closure: $G^{(k)} = \overline{F^{(k)}}$, i.e. $G^{(k)}$ is obtained by completing every connected components of $F^{(k)}$

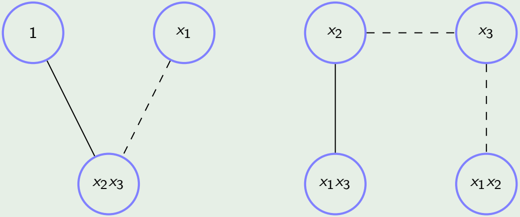
Support-extension

Example

Consider the following graph $G(V, E)$ with

$$V = \{1, x_1, x_2, x_3, x_2x_3, x_1x_3, x_1x_2\} \text{ and } E = \{\{1, x_2x_3\}, \{x_2, x_1x_3\}\}.$$

Figure: The support-extension of G



Block-closure

Example

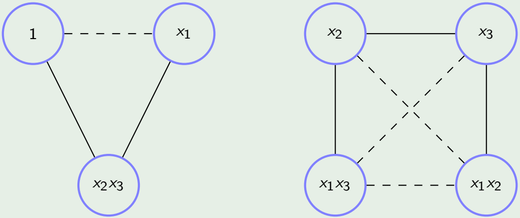
Consider the following graph $G(V, E)$ with

$$V = \{1, x_1, x_2, x_3, x_2x_3, x_1x_3, x_1x_2\}$$

and

$$E = \{\{1, x_2x_3\}, \{x_2, x_1x_3\}, \{x_1, x_2x_3\}, \{x_2, x_3\}, \{x_3, x_1x_2\}\}.$$

Figure: The block-closure of G



Term sparsity adapted moment-SOS hierarchies

Let $\mathscr{B}_1^{(k)}, \ldots, \mathscr{B}_{t_k}^{(k)}$ be the connected components of $G^{(k)}$. For each $k \geq 1$, consider

$$(\mathrm{Q}^k): \quad \begin{array}{rll} \theta^k := & \inf & L_{\mathbf{y}}(f) \\ & \text{s.t.} & M_{\mathscr{B}_i^{(k)}}(\mathbf{y}) \succeq 0, \quad i = 1, \ldots, t_k \\ & & y_{\mathbf{0}} = 1. \end{array}$$

We have

$$\theta^1 \leq \theta^2 \leq \cdots \leq \theta^*.$$

We call $(\mathrm{Q}^k), k = 1, 2, \ldots$ the TSSOS hierarchy for (Q) and k the sparse order.

A two-level hierarchy of lower bounds

This procedure easily extends to the constrained case. Consequently, we obtain a two-level hierarchy of lower bounds for the optimum θ^* of (Q): $(r = \max\{\deg(f)/2, d_1, \ldots, d_m\})$

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

Finite convergence

Theorem

Fixing a relaxation order d, the sequence $(\theta_d^{(k)})_{k\geq 1}$ converges to θ_d in finitely many steps.

Sign symmetries

Definition

Given a finite set $\mathscr{A} \subseteq \mathbb{N}^n$, the sign symmetries of $\mathscr{A}$ are defined by all vectors $\mathbf{r} \in \mathbb{Z}_2^n$ such that $\mathbf{r}^T \boldsymbol{\alpha} \equiv 0 \pmod 2$ for all $\boldsymbol{\alpha} \in \mathscr{A}$.

Sign symmetries

Definition

Given a finite set $\mathscr{A} \subseteq \mathbb{N}^n$, the sign symmetries of $\mathscr{A}$ are defined by all vectors $\mathbf{r} \in \mathbb{Z}_2^n$ such that $\mathbf{r}^T \boldsymbol{\alpha} \equiv 0 \pmod 2$ for all $\boldsymbol{\alpha} \in \mathscr{A}$.

Example

Let $\mathscr{A} = \{\begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}\}$. The sign symmetries of $\mathscr{A}$ are $\mathbf{r}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mathbf{r}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Sign symmetries

Assume $\mathscr{A} = \{\boldsymbol{\alpha} \in \mathbb{N}^n \mid \mathbf{x}^{\boldsymbol{\alpha}} \in \mathrm{supp}(f) \cup \bigcup_{j=1}^m \mathrm{supp}(g_j)\}$.

The set of sign symmetries $R := [\mathbf{r}_1, \ldots, \mathbf{r}_s]$ of $\mathscr{A}$ induces a partition of $V_d(\mathbf{x})$, $V_{d-d_j}(\mathbf{x})$:
$\mathbf{x}^{\boldsymbol{\alpha}}, \mathbf{x}^{\boldsymbol{\beta}}$ belong to the same block $\iff R^T(\boldsymbol{\alpha} + \boldsymbol{\beta}) \equiv 0 \pmod 2$.

Sign symmetries

Assume $\mathscr{A} = \{\boldsymbol{\alpha} \in \mathbb{N}^n \mid \mathbf{x}^{\boldsymbol{\alpha}} \in \mathrm{supp}(f) \cup \bigcup_{j=1}^{m} \mathrm{supp}(g_j)\}$.

The set of sign symmetries $R := [\mathbf{r}_1, \ldots, \mathbf{r}_s]$ of $\mathscr{A}$ induces a partition of $V_d(\mathbf{x})$, $V_{d-d_j}(\mathbf{x})$:
$\mathbf{x}^{\boldsymbol{\alpha}}, \mathbf{x}^{\boldsymbol{\beta}}$ belong to the same block $\iff R^T(\boldsymbol{\alpha} + \boldsymbol{\beta}) \equiv 0 \pmod 2$.

Theorem

Fixing a relaxation order d, the partition of monomial bases $V_d(\mathbf{x})$, $V_{d-d_j}(\mathbf{x})$ at the final step of the TSSOS hierarchy is the one induced by the sign symmetries of the above $\mathscr{A}$.

Extensions

- Replacing block-closure by chordal-extension;
- Exploiting correlative sparsity and term sparsity simultaneously;
- Exploiting quotient structure and term sparsity simultaneously;
- Extending to complex polynomial optimization;
- Extending to noncommutative polynomial optimization;
- ......

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; 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; 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	-	-	-

AC-OPF problems

Table: The results for AC-OPF problems; the symbol "-" indicates out of memory

n	m	CS ($d=2$)				CS+TS ($d=2$)			
		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%

Eigenvalue minimization for the noncommutative generalized Rosenbrock function

Table: The eigenvalue minimization for the noncommutative generalized Rosenbrock function over $\mathcal{D}$, where $\mathcal{D}$ is defined by $\{1 - X_1^2, \ldots, 1 - X_n^2, X_1 - 1/3, \ldots, X_n - 1/3\}$; the symbol "-" indicates out of memory

n	CS+TS ($d = 2$)			Dense ($d = 2$)		
	mb	opt	time (s)	mb	opt	time (s)
20	3	1.0000	0.14	-	-	-
40	3	1.0000	0.22	-	-	-
60	3	0.9999	0.28	-	-	-
80	3	0.9999	0.35	-	-	-
100	3	0.9999	0.46	-	-	-
200	3	0.9999	0.89	-	-	-
400	3	1.0000	2.40	-	-	-
600	3	1.0000	4.47	-	-	-
800	3	1.0000	6.95	-	-	-
1000	3	0.9999	10.2	-	-	-
2000	3	0.9999	37.2	-	-	-
3000	3	0.9999	87.2	-	-	-
4000	3	0.9998	145	-	-	-

Conclusions and outlooks

- The concept of term sparsity patterns opens a new window to exploit sparsity at the term level for polynomial optimization, in contrast to the usual correlative sparsity pattern;
- The TSSOS hierarchy is a powerful tool to handle large-scale polynomial optimization problems;
- One can exploit term sparsity for generalized moment problems (more general than polynomial optimization);
- Fruitful potential applications: optimal power flow, computer vision, control, quantum information, tensor decomposition ...

Main references

- Jie Wang, Victor Magron and Jean B. Lasserre, *TSSOS: A Moment-SOS hierarchy that exploits term sparsity*, arXiv:1912.08899 (2019).
- Jie Wang, Victor Magron and Jean B. Lasserre, *Chordal-TSSOS: a moment-SOS hierarchy that exploits term sparsity with chordal extension*, arXiv:2003.03210 (2020).
- 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, Martina Maggio and Victor Magron, *SparseJSR: A Fast Algorithm to Compute Joint Spectral Radius via Sparse SOS Decompositions*, arXiv:2008.11441 (2020).
- Jared Miller, Jie Wang, Mario Sznaier and Octavia Camps, *Model Fitting by Semialgebraic Clustering*, 2020.
- Jie Wang and Victor Magron, Exploiting Term-Sparsity in Noncommutative Polynomial Optimization, upcoming (2020).
- `TSSOS`: https://github.com/wangjie212/TSSOS
- `NCTSSOS`: https://github.com/wangjie212/NCTSSOS
- `SparseJSR`: https://github.com/wangjie212/SparseJSR