Exploiting Sparsity in Noncommutative Polynomial Optimization

Jie Wang
Joint work with Victor Magron

21 June, ECM 2021

Content

1. Noncommutative polynomial optimization and the moment-SOHS hierarchy
2. Exploiting sparsity for the moment-SOHS hierarchy
3. Numerical experiments

Polynomials in noncommutating variables

- $\underline{X} = (X_1, \ldots, X_n)$: a tuple of noncommutating variables (letters)
- noncommutative (nc) polynomial: $f = \sum_{w \in \langle \underline{X} \rangle} a_w w$, $a_w \in \mathbb{R}$
- $\mathbb{R}\langle \underline{X} \rangle$: the ring of nc polynomials in $\underline{X}$
- involution $\star$: fixes $\mathbb{R} \cup \{X_1, \ldots, X_n\}$ and reverses words (monomials)
- $\mathrm{Sym}\,\mathbb{R}\langle \underline{X} \rangle := \{f \in \mathbb{R}\langle \underline{X} \rangle \mid f^\star = f\}$
- sum of Hermitian squares (SOHS): $f = g_1^\star g_1 + g_2^\star g_2 + \ldots + g_r^\star g_r$

Positivity over a semialgebraic set

Given $S = \{g_1, \ldots, g_m\} \subseteq \operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle$, the semialgebraic set $\mathcal{D}_S$ is

$$\mathcal{D}_S := \bigcup_{r \in \mathbb{N} \setminus \{0\}} \{\underline{A} = (A_1, \ldots, A_n) \in (\mathbb{S}^r)^n \mid g_j(\underline{A}) \succeq 0, j \in [m]\}.$$

The operator semialgebraic set $\mathcal{D}_S^\infty$ is the set of all bounded self-adjoint operators $\underline{A}$ on a Hilbert space endowed with a scalar product $\langle \cdot, \cdot \rangle$ making $g_j(\underline{A})$ a PSD operator, for all $g_j \in S$.

Theorem (Helton and McCullough, 2002)

Let $\{f\} \cup S \subseteq \operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle$ and assume that $\mathcal{M}_S$ is Archimedean. If $f(A) \succ 0$ for all $A \in \mathcal{D}_S^\infty$, then $f \in \mathcal{M}_S$.

Eigenvalue minimization and trace minimization

Given $f \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X} \rangle$ and $S = \{g_1, \ldots, g_m\} \subseteq \mathrm{Sym}\,\mathbb{R}\langle \underline{X} \rangle$, the **eigenvalue minimization problem** for f over the operator semialgebraic set $\mathcal{D}_S^\infty$ is defined by

$$\lambda_{\min}(f, S) := \inf\{\langle f(\underline{A})\boldsymbol{v}, \boldsymbol{v}\rangle : \underline{A} \in \mathcal{D}_S^\infty, ||\boldsymbol{v}|| = 1\},$$

and the **trace minimization problem** for f over the semialgebraic set $\mathcal{D}_S$ is defined by

$$\mathrm{tr}_{\min}(f, S) := \inf\{\mathrm{tr}\, f(\underline{A}) : \underline{A} \in \mathcal{D}_S\}.$$

(For $A = [a_{ij}] \in \mathbb{S}^r$, $\mathrm{tr}\, A := \frac{1}{r}\sum_{i=1}^r a_{ii}$.)

Moment matrix and localizing matrix

For $d \in \mathbb{N}$, let $\mathbf{W}_d$ be the column vector of all words of degree at most d arranged w.r.t. the lexicographic order.

$\triangleright$ The d-th order moment matrix $M_d(\mathbf{y})$ is defined by

$$[M_d(\mathbf{y})]_{uv} := y_{u^\star v}, \quad \forall u, v \in \mathbf{W}_d.$$

$\triangleright$ Given $g = \sum_{w \in \mathrm{supp}(g)} b_w w \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X} \rangle$, the d-th order localizing matrix $M_d(g\mathbf{y})$ is defined by

$$[M_d(g\mathbf{y})]_{uv} := \sum_{w \in \mathrm{supp}(g)} b_w y_{u^\star w v}, \quad \forall u, v \in \mathbf{W}_d.$$

The moment-SOHS hierarchy for eigenvalue minimization

The d-th order moment relaxation:

$$\begin{aligned}\lambda_d(f,S) := \quad \inf \quad & L_{\mathbf{y}}(f) \\ \text{s.t.} \quad & M_d(\mathbf{y}) \succeq 0, \\ & M_{d-d_j}(g_j\mathbf{y}) \succeq 0, \quad j = 1, \ldots, m, \\ & y_1 = 1,\end{aligned}$$

with the dual SOHS relaxation:

$$\begin{aligned}\lambda_d(f,S)^* := \quad \sup \quad & \lambda \\ \text{s.t.} \quad & f - \lambda \in \mathcal{M}_{S,2d}.\end{aligned}$$

$(d_j = \lceil \deg(g_j)/2 \rceil, j = 1, \ldots, m)$

The moment-SOHS hierarchy for eigenvalue minimization

$$
\begin{array}{ccccccc}
 & & & \lambda_{\min}(f,S) & & & \\
 & & \rotatebox{-45}{$\leq$} & & \rotatebox{45}{$\geq$} & & \\
 & \vdots & & & & \vdots & \\
 & \geq & & & & \geq & \\
\text{Moment} & \lambda_d(f,S) & & \text{``}=\text{''} & & \lambda_d(f,S)^* & \text{SOHS} \\
 & \geq & & & & \geq & \\
 & \vdots & & & & \vdots & \\
 & \geq & & & & \geq & \\
 & \lambda_{\underline{d}}(f,S) & & \text{``}=\text{''} & & \lambda_{\underline{d}}(f,S)^* &
\end{array}
$$

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

The moment-SOHS hierarchy for trace minimization

The d-th order moment relaxation:

$$\begin{array}{rll} \mu_d(f,S) := & \inf & L_{\mathbf{y}}(f) \\ & \text{s.t.} & M_d(\mathbf{y}) \succeq 0, \\ & & M_{d-d_j}(g_j\mathbf{y}) \succeq 0, \quad j \in [m], \\ & & M_d(\mathbf{y})_{uv} = M_d(\mathbf{y})_{wz}, \quad \text{for all } u^\star v \overset{\text{cyc}}{\sim} w^\star z, \\ & & y_1 = 1, \end{array}$$

with the dual SOHS relaxation:

$$\begin{array}{rll} \mu_d(f,S)^* := & \sup & \lambda \\ & \text{s.t.} & f - \lambda \in_{\text{cyc}} \mathcal{M}_{S,2d}. \end{array}$$

Global convergence

Under Archimedean's condition: there exists $N > 0$ s.t. $N - ||\underline{X}||^2 \in \mathcal{M}_S$, we have

- $\lambda_d(f, S) \uparrow \lambda_{\min}(f, S)$ and $\lambda_d(f, S)^* \uparrow \lambda_{\min}(f, S)$ as $d \to \infty$;
- We can verify global optimality by the so-called rank condition (flat extension);
- We can easily extract minimizers when the rank condition is satisfied.

Global convergence

Under Archimedean's condition: there exists $N > 0$ s.t. $N - ||\underline{X}||^2 \in \mathcal{M}_S$, we have

- $\lambda_d(f, S) \uparrow \lambda_{\min}(f, S)$ and $\lambda_d(f, S)^* \uparrow \lambda_{\min}(f, S)$ as $d \to \infty$;
- We can verify global optimality by the so-called rank condition (flat extension);
- We can easily extract minimizers when the rank condition is satisfied.

Key Message: The moment-SOHS 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 SOHS problem for eigenvalue minimization) at relaxation order d:

- maximal size of PSD matrices: $\frac{n^{d+1}-1}{n-1}$
- #equality constraint: $\frac{n^{2d+1}-1}{n-1}$

In view of the current state of SDP solvers (e.g. `Mosek`), solvable problems are limited to modest size.

Scalability issue

The size of SDP (considering the SOHS problem for eigenvalue minimization) at relaxation order d:

- maximal size of PSD matrices: $\frac{n^{d+1}-1}{n-1}$
- #equality constraint: $\frac{n^{2d+1}-1}{n-1}$

In view of the current state of SDP solvers (e.g. `Mosek`), solvable problems are limited to modest size.

Exploiting structure:

- symmetry
- sparsity

Correlative sparsity (Klep, Magron and Povh, 2021)

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

Correlative sparsity pattern (csp) graph $G^{\mathrm{csp}}(V, E)$:
$V := \{X_1, \ldots, X_n\}$
$\{X_i, X_j\} \in E \iff X_i, X_j$ appear in the same term of f or appear in the same constraint polynomial g_k

Correlative sparsity (Klep, Magron and Povh, 2021)

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:

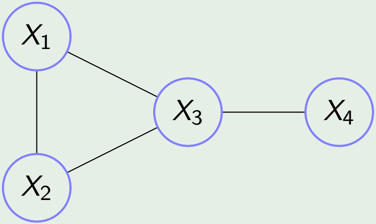
$$I_k \longmapsto M_d(\mathbf{y}, I_k), M_{d-d_j}(g_j\mathbf{y}, I_k)$$

Correlative sparsity

Example

Consider $f = X_1^4 + X_1X_2^2 + X_2^2X_1 + X_2X_3 + X_3X_2 + X_3^2X_4^2 + X_4^2X_3^2$ and $g_1 = 1 - X_1^2 - X_2^2 - X_3^2$, $g_2 = 1 - X_3X_4 - X_4X_3$.

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



Two maximal cliques: $\{X_1, X_2, X_3\}$ and $\{X_3, X_4\}$

The correlative sparsity adapted moment-SOHS hierarchy

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

Term sparsity (Wang and Magron, 2021)

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

Term sparsity (Wang and Magron, 2021)

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

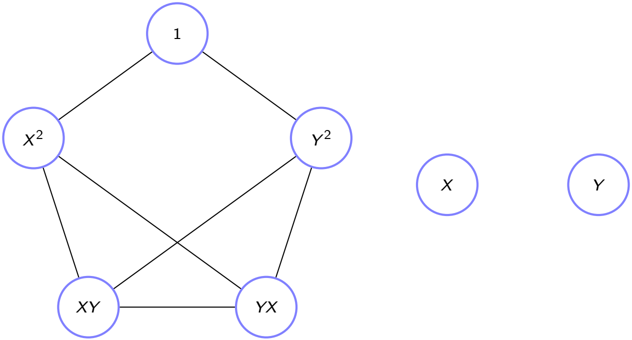
Term sparsity pattern (tsp) graph $G^{\rm tsp}(V, E)$ (with relaxation order d):
$V := \mathbf{W}_d$
$\{u, v\} \in E \Longleftrightarrow u^\star v \in \mathrm{supp}(f) \cup \bigcup_{j=1}^m \mathrm{supp}(g_j) \cup \{w^\star w \mid w \in \mathbf{W}_d\}$

(For $f = \sum_{w \in \langle \underline{X} \rangle} a_w w \in \mathbb{R}\langle \underline{X} \rangle$, $\mathrm{supp}(f) := \{w \mid a_w \neq 0\}$)

Term sparsity

Let
$f = 2 - X^2 + XY^2X - Y^2 + XYXY + YXYX + X^3Y + YX^3 + XY^3 + Y^3X$
and $S = \{1 - X^2, 1 - Y^2\}$.

Figure: The tsp graph for f and S with $d = 2$



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

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

In the moment relaxation,

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

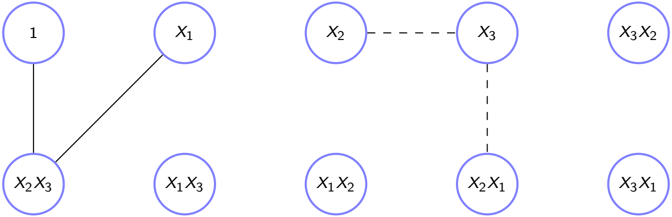
Similarly for the localizing matrices $M_{d-d_j}(\mathbf{y}), j = 1, \ldots, m$.

Extending to an iterative procedure

By iteratively performing support extension and chordal extension:

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

Figure: Support extension ($w_1^\star w_2 = w_3^\star w_4$ and $\{w_1, w_2\} \in E \Rightarrow \{w_3, w_4\} \in E$)



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

$$\begin{array}{rll} \lambda_d^{(s)}(f,S) := & \inf & L_{\mathbf{y}}(f) \\ & \text{s.t.} & M_{C_{0,i}^{(s)}}(\mathbf{y}) \succeq 0, \quad i = 1,\ldots,t_{0,s}, \\ & & M_{C_{j,i}^{(s)}}(g_j\mathbf{y}) \succeq 0, \quad i = 1,\ldots,t_{j,s}, j = 1,\ldots,m, \\ & & y_1 = 1. \end{array}$$

We obtain

$$\lambda_d^{(1)}(f,S) \leq \lambda_d^{(2)}(f,S) \leq \cdots \leq \lambda_d(f,S).$$

A two-level hierarchy of lower bounds

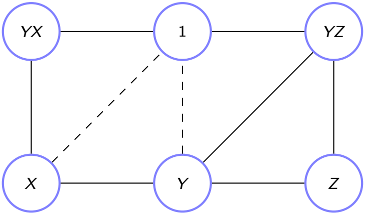
Furthermore, we obtain a two-level hierarchy of lower bounds for $\lambda_{\min}(f,S)$: ($\underline{d} := \max\{\deg(f)/2, d_1, \ldots, d_m\}$)

$$\begin{array}{ccccccc}
\lambda_{\underline{d}}^{(1)}(f,S) & \leq & \lambda_{\underline{d}}^{(2)}(f,S) & \leq & \cdots & \leq & \lambda_{\underline{d}}(f,S) \\
\wedge| & & \wedge| & & & & \wedge| \\
\lambda_{\underline{d}+1}^{(1)}(f,S) & \leq & \lambda_{\underline{d}+1}^{(2)}(f,S) & \leq & \cdots & \leq & \lambda_{\underline{d}+1}(f,S) \\
\wedge| & & \wedge| & & & & \wedge| \\
\vdots & & \vdots & & \vdots & & \vdots \\
\wedge| & & \wedge| & & & & \wedge| \\
\lambda_{d}^{(1)}(f,S) & \leq & \lambda_{d}^{(2)}(f,S) & \leq & \cdots & \leq & \lambda_{d}(f,S) \\
\wedge| & & \wedge| & & & & \wedge| \\
\vdots & & \vdots & & \vdots & & \vdots
\end{array}$$

Example

Consider the nc polynomial $f = X^2 - XY - YX + 3Y^2 - 2XYX + 2XY^2X - YZ - ZY + 6Z^2 + 9X^2Y + 9Z^2Y - 54ZYZ + 142ZY^2Z$. The monomial basis given by the Newton chip method is $\{1, X, Y, Z, YX, YZ\}$. Solving the sparse moment relaxation with $s = 1$, we obtain $\lambda^{(1)}(f) \approx -0.00355$ while $\lambda_{\min}(f) = 0$.

Figure: The tsp graph and a chordal extension



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 $\lambda_{\min}(f, S)$.

Eigenvalue minimization for the nc Broyden banded function

Software: `NCTSSOS`, SDP solver : `Mosek`

Table: Eigenvalue minimization for the nc Broyden banded function with $d = 3, s = 1$; the symbol "-" indicates out of memory.

n	Sparse			Dense		
	mb	opt	time (s)	mb	opt	time (s)
20	15	0	0.34	61	0	1.42
40	15	0	0.77	121	0	34.9
60	15	0	0.97	181	0	367
80	15	0	1.20	-	-	-
100	15	0	1.57	-	-	-
200	15	0	3.14	-	-	-
300	15	0	5.25	-	-	-
400	15	0	7.11	-	-	-
500	15	0	9.42	-	-	-
600	15	0	12.9	-	-	-
700	15	0	15.6	-	-	-
800	15	0	18.5	-	-	-
900	15	0	22.3	-	-	-
1000	15	0	26.2	-	-	-

Eigenvalue minimization for the nc generalized Rosenbrock function

Table: Eigenvalue minimization for the nc generalized Rosenbrock function over $\mathcal{D}$ with $d = 2, s = 1$, 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	Sparse			Dense		
	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	-	-	-

Eigenvalue minimization for randomly generated examples

Table: Eigenvalue minimization for random nc quartic polynomials over multi-balls with $d = 2$; the symbol "-" indicates out of memory.

n	CS+TS				CS			Dense		
	s	mb	opt	time (s)	mb	opt	time (s)	mb	opt	time (s)
505	1	21	-15.91	3.26	241	-	-	-	-	-
	2	21	-15.42	7.49						
	3	21	-15.31	10.6						
1005	1	25	-32.58	9.71	241	-	-	-	-	-
	2	25	-31.91	24.5						
	3	25	-31.71	40.9						
1505	1	26	-48.57	18.9	241	-	-	-	-	-
	2	26	-47.00	47.0						
	3	26	-46.71	90.0						
2005	1	25	-63.58	33.7	241	-	-	-	-	-
	2	25	-62.05	85.8						
	3	25	-61.76	149						
2505	1	23	-81.07	52.9	241	-	-	-	-	-
	2	23	-78.75	134						
	3	23	-78.21	263						
3005	1	23	-95.73	74.8	241	-	-	-	-	-
	2	23	-93.13	212						
	3	23	-92.71	396						
3505	1	24	-111.2	93.4	241	-	-	-	-	-
	2	24	-108.3	258						
	3	24	-107.8	531						
4005	1	25	-131.1	122	241	-	-	-	-	-
	2	25	-127.5	375						
	3	25	-126.8	687						

Trace minimization for the nc Broyden tridiagonal function

Table: Trace minimization for the nc Broyden tridiagonal function over $\mathcal{D}$ with $d = 2, s = 1$, 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	Sparse			Dense		
	mb	opt	time (s)	mb	opt	time (s)
20	6	1.1805	0.27	-	-	-
40	6	1.1828	0.53	-	-	-
60	6	1.1828	0.68	-	-	-
80	6	1.1828	0.82	-	-	-
100	6	1.1828	1.07	-	-	-
200	6	1.1828	2.45	-	-	-
400	6	1.1828	6.18	-	-	-
600	6	1.1828	12.2	-	-	-
800	6	1.1828	20.1	-	-	-
1000	6	1.1828	28.6	-	-	-
2000	6	1.1828	104	-	-	-
3000	6	1.1828	204	-	-	-
4000	6	1.1828	363	-	-	-

Trace minimization for randomly generated examples

Table: Trace minimization for random nc quartic polynomials over multi-balls with $d = 2$; the symbol "-" indicates out of memory.

n	CS+TS				CS			Dense		
	s	mb	opt	time (s)	mb	opt	time (s)	mb	opt	time (s)
505	1	16	-4.997	4.94	241	-	-	-	-	-
	2	17	-4.983	7.40						
	3	17	-4.975	7.66						
1005	1	16	-10.14	14.2	241	-	-	-	-	-
	2	17	-10.11	21.7						
	3	17	-10.11	22.6						
1505	1	16	-15.72	25.2	241	-	-	-	-	-
	2	17	-15.68	39.8						
	3	17	-15.67	41.0						
2005	1	16	-20.45	40.9	241	-	-	-	-	-
	2	17	-20.41	67.9						
	3	17	-20.40	73.8						
2505	1	16	-25.95	63.1	241	-	-	-	-	-
	2	17	-25.90	95.6						
	3	18	-25.89	101						
3005	1	16	-31.09	93.5	241	-	-	-	-	-
	2	17	-31.03	152						
	3	18	-31.02	157						
3505	1	16	-35.99	119	241	-	-	-	-	-
	2	17	-35.93	198						
	3	18	-35.92	216						
4005	1	16	-41.80	145	241	-	-	-	-	-
	2	17	-41.72	248						
	3	18	-41.70	264						

Conclusion and outlook

1. When appropriate sparsity patterns are accessible, we can significantly improve the scalability of the moment-SOHS hierarchy;
2. Possible extensions to other situations also relying on the moment-SOHS hierarchy, e.g., problems involving trace polynomials;
3. Potential applications in quantum information, condensed matter physics and other fields.

Main references

- Jie Wang, Victor Magron and Jean B. Lasserre, *TSSOS: A Moment-SOS hierarchy that exploits term sparsity*, SIAM Optimization, 2020.
- Jie Wang, Victor Magron and Jean B. Lasserre, *Chordal-TSSOS: a moment-SOS hierarchy that exploits term sparsity with chordal extension*, SIAM Optimization, 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 and Victor Magron, Exploiting Term Sparsity in Noncommutative Polynomial Optimization, arXiv:2010.06956, 2020.
- `TSSOS`: https://github.com/wangjie212/TSSOS
- `NCTSSOS`: https://github.com/wangjie212/NCTSSOS

Thanks for your attention!