Computational Optimization and Applications (2021) 80:483–521
https://doi.org/10.1007/s10589-021-00301-7

Exploiting term sparsity in noncommutative polynomial optimization

Jie Wang[1] · **Victor Magron**[1]

Received: 15 October 2020 / Accepted: 7 July 2021 / Published online: 27 July 2021

Abstract
We provide a new hierarchy of semidefinite programming relaxations, called *NCTSSOS*, to solve large-scale sparse noncommutative polynomial optimization problems. This hierarchy features the exploitation of *term sparsity* hidden in the input data for eigenvalue and trace optimization problems. NCTSSOS complements the recent work that exploits *correlative sparsity* for noncommutative optimization problems by Klep et al. (MP, 2021), and is the noncommutative analogue of the TSSOS framework by Wang et al. (SIAMJO 31: 114–141, 2021, SIAMJO 31: 30–58, 2021). We also propose an extension exploiting simultaneously correlative and term sparsity, as done previously in the commutative case (Wang in CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization, 2020). Under certain conditions, we prove that the optima of the NCTSSOS hierarchy converge to the optimum of the corresponding dense semidefinite programming relaxation. We illustrate the efficiency and scalability of NCTSSOS by solving eigenvalue/trace optimization problems from the literature as well as randomly generated examples involving up to several thousand variables.

Keywords Term sparsity · Noncommutative polynomial optimization · Semidefinite relaxations · Sum of hermitian squares · Eigenvalue optimization · Trace optimization

Mathematics Subject Classification 47N10 · 90C22 · 90C23 · 12D15 · 14P10

✉ Victor Magron
jwang@laas.fr

Jie Wang
vmagron@laas.fr

[1] Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS), Toulouse, France

1 Introduction

A polynomial optimization problem (POP) consists of minimizing a polynomial over a basic closed *semialgebraic set*, namely an intersection of finitely many polynomial superlevel sets. Even if solving a POP is NP-hard in general [25], one can rely on the so-called "moment-sums of squares (SOS) hierarchy", also referred to as "Lasserre's hierarchy" [23] to compute a sequence of lower bounds for the POP. Each lower bound in the sequence is obtained by solving a semidefinite program (SDP) [2]. Thanks to Putinar's Positivstellensatz [35], if the quadratic module generated by the polynomials describing the semialgebraic set is *Archimedean*, the sequence of these SDP lower bounds converges from below to the global optimum of the POP.

Although most POPs involve commuting variables, we are also interested in noncommutative POPs (NCPOP), i.e., POPs involving noncommuting variables (e.g. matrices, operators on a Hilbert space). The applications of NCPOPs include control theory and linear systems in engineering [37], quantum theory and quantum information science [16, 30, 32], matrix factorization ranks [15], machine learning [50, 51] and so on, and new applications are emerging.

The noncommutative (nc) analogue of Lasserre's hierarchy [12, 31, 33], often called the "Navascués-Pironio-Acín (NPA) hierarchy", or "moment-sums of hermitian squares (SOHS) hierarchy", allows one to compute arbitrarily close lower bounds of the minimal eigenvalue of an nc polynomial over an nc semialgebraic set. In the same spirit, one can also obtain a hierarchy of SDP relaxations to approximate as closely as desired the minimal trace of an nc polynomial over an nc semialgebraic set [9]. We also refer the interested reader to [22] for the case of more general trace polynomials, i.e., polynomials in noncommuting variables and traces of their products.

From the view of applications, the common bottleneck of the Lasserre/NPA hierarchy is that the involved sequence of SDP relaxations becomes intractable very quickly as the number of variables n or the relaxation order d increases. In the commutative setting, the matrices involved at relaxation order d is of size $\binom{n+d}{d}$; in the nc setting, the size of matrices involved at relaxation order d is even larger. It is already hard to solve such an SDP for $n \simeq 30$ ($n \simeq 20$ in the nc setting) and $d = 2$ on a modern standard laptop (at least when one relies on interior-point solvers).

1.1 Remedies by exploiting sparsity for (NC)POP

In certain situations, the SDP relaxations arising from POPs can be solved with adequate first-order methods rather than with costly interior-point algorithms; see e.g., [28, 49], where the authors exploit the constant trace property of the matrices in SDP relaxations of quadratically constrained quadratic programs or combinatorial optimization problems. In any case, it is worth finding remedies in view of the sparsity of (NC)POPs to prevent from the computational blow-up of the Lasserre/NPA hierarchy, by decreasing the sizes of the matrices involved in the SDP relaxations.

The first remedy is to partition the input variables into cliques when the polynomials involved in the objective function and the constraints fulfill a so-called *correlative sparsity pattern*. The resulting moment-SOS hierarchy is obtained by assembling sparse SDP matrices in terms of these cliques of variables [41, 42]. Under certain conditions, the lower bounds given by this hierarchy still converge to the global optimum of the original problem [24]. When the sizes of these cliques of variables are small enough (e.g., less than 10 in [41] or less than 20 in [20]), one can significantly improve the scalability of Lasserre's hierarchy to handle problems with a large number of variables. For instance, by exploiting correlative sparsity, one can compute roundoff error bounds [26, 27] with up to several hundred variables, and solve optimal power flow problems [20] or deep learning problems [13] with up to thousands of variables. Several extensions have been investigated, including volume computation of sparse semialgebraic sets [38], or minimization of rational functions [8, 29]. Recently, Klep, the second author and Povh designed an nc analogue of this sparsity adapted hierarchy for both eigenvalue and trace minimization problems [21]. Nevertheless, when the sizes of variable cliques provided by correlative sparsity are relatively big (say ≥ 20 in the commutative setting and ≥ 10 in the nc setting), or when the correlative sparsity pattern is even fully dense, one might face again the same issue of untractable SDPs.

Another complementary remedy consists of exploiting *term sparsity*. For unconstrained problems, it means that the objective function involves few terms (monomials or words). One can then reduce the size of the associated SDP matrix by computing a smaller monomial basis via the *Newton polytope method* [36]. The nc analogue for this method is the *Newton chip method* [11, §2.3] in the context of eigenvalue optimization and the *tracial Newton polytope* [11, §3.3] in the context of trace optimization.

Besides obtaining a smaller monomial basis, in both unconstrained and constrained cases, one can rely on a term-sparsity adapted moment-SOS hierarchy (called TSSOS), following the line of research recently pursued by the two authors, Lasserre and Mai in [44, 46, 47]. The core idea of TSSOS is partitioning the monomial bases used to construct SDP relaxations into blocks in view of the correlations between monomials and then building SDP matrices to comply with this block structure. More precisely, one first define the *term sparsity pattern (tsp) graph* associated with a POP whose nodes are monomials from the monomial basis. Two nodes of the tsp graph are connected via an edge if and only if the product of the corresponding monomials belongs to the supports of the polynomials involved in the POP or is a monomial square. TSSOS is based on an iterative procedure, whose input is the tsp graph of the POP. Each iteration consists of two steps: first one performs a support extension operation on the graph and successively one performs a chordal extension operation on the graph ("maximal" chordal extensions are used in [47] while approximately minimum chordal extensions are used in [46]). At each iteration, one can construct an SDP relaxation with matrices of sparsity pattern represented by the corresponding graph. In doing so, TSSOS provides us with a two-level moment-SOS hierarchy involving sparse SDP matrices. TSSOS can be further combined with correlative sparsity, which allows one to solve large-scale POPs with

several thousand variables and constraints [48]. Apart from (commutative) POPs, the idea of TSSOS can be also used to develop more efficient SOS-based algorithms for other problems, e.g. the approximation of joint spectral radius [45].

Contributions Motivated by the performance of TSSOS for commutative POPs, we develop an nc analogue of the TSSOS framework (called NCTSSOS) in this paper, in order to handle large-scale eigenvalue/trace optimization problems with sparse input data.

First, we extend in Sect. 3 the notion of term sparsity patterns to unconstrained eigenvalue optimization problems. We show how to build the tsp graph and derive a two-step iterative procedure to enlarge the graph via the support extension operation and the chordal extension operation. Based on this, we then give the NCTSSOS hierarchy. The generalization to constrained eigenvalue optimization problems is provided in Sect. 3.2. Under certain conditions, we prove that the optima of the NCTSSOS hierarchy converge to the optimum of the corresponding dense relaxation. In Sect. 4, we show how to benefit simultaneously from both correlative and term sparsity, to obtain an nc variant of the so-called "CS-TSSOS" hierarchy [48]. Section 5 is dedicated to trace optimization. For both unconstrained and constrained trace optimization problems, we provide a term sparsity (and combined with correlative sparsity) adapted hierarchy of SDP relaxations. In Sect. 6, we demonstrate the computational efficiency, scalability and accuracy of the NCTSSOS hierarchy by various numerical examples involving up to several thousand variables.

The algorithmic framework of the NCTSSOS hierarchy has been released as an open-source Julia [4] package, also called `NCTSSOS`, which is available online and comes together with a documentation.[1]

Our term sparsity (and combined with correlative sparsity) adapted moment-SOHS hierarchies appear in a similar manner as the ones obtained for the commutative case [46–48]. We believe that it is of interest for researchers using non-commutative optimization tools to have a self-contained paper stating explicitly the construction of tsp graphs, support/chordal extension operations, as well as the different term sparsity (and combined with correlative sparsity) adapted SDP formulations, either for eigenvalue or trace optimization. While the overall strategy to obtain tsp graphs for eigenvalue optimization is very similar to the commutative case, it is less straightforward for trace optimization, where it is mandatory to introduce the cyclic analog of the tsp graph and the support extension operation. Furthermore, we would like to emphasize that the main contribution of this paper is to show a significant quantitative improvement with respect to the previous results obtained for various nc eigenvalue/trace optimization problems. We hope that these results will convince researchers in related fields, including quantum information physicists relying on the NPA hierarchy, about the potential impact that NCTSSOS could have on solving their problems more efficiently.

[1] https://github.com/wangjie212/NCTSSOS.

2 Notation and preliminaries

In this section, we recall some notations, definitions and basic results that will be used in the rest of this paper.

2.1 Noncommutative polynomials

For a positive integer r, let us denote by $\mathbb{S}_r$ (resp. $\mathbb{S}_r^+$) the space of all symmetric (resp. positive semidefinite (PSD)) matrices of size r, and by $\mathbb{S}_r^n$ the set of n-tuples $\underline{A} = (A_1, \ldots, A_n)$ of symmetric matrices A_i of size r. For matrices $A, B \in \mathbb{S}_r$ (resp. vectors $\mathbf{u}, \boldsymbol{v} \in \mathbb{R}^r$), let $\langle A, B\rangle \in \mathbb{R}$ (resp. $\langle \mathbf{u}, \boldsymbol{v}\rangle \in \mathbb{R}$) be the trace inner-product, defined by $\langle A, B\rangle = \mathrm{Tr}(A^T B)$ (resp. $\langle \mathbf{u}, \boldsymbol{v}\rangle = \mathbf{u}^T \boldsymbol{v}$) and let $A \circ B \in \mathbb{S}_r$ denote the Hadamard, or entrywise, product of A and B, defined by $[A \circ B]_{ij} = A_{ij} B_{ij}$. For a fixed $n \in \mathbb{N}\backslash\{0\}$, let $\underline{X} = (X_1, \ldots, X_n)$ be a tuple of letters and consider the set of all possible words of finite length in $\underline{X}$ which is denoted by $\langle \underline{X}\rangle$. The empty word is denoted by 1. We denote by $\mathbb{R}\langle \underline{X}\rangle$ the ring of real polynomials in the noncommutating variables $\underline{X}$. An element f in $\mathbb{R}\langle \underline{X}\rangle$ can be written as $f = \sum_{w \in \langle \underline{X}\rangle} a_w w$, $a_w \in \mathbb{R}$, which is called a *noncommutative polynomial* (*nc polynomial* for short). The *support* of f is defined by $\mathrm{supp}(f) := \{w \in \langle \underline{X}\rangle \mid a_w \neq 0\}$ and the *degree* of f, denoted by $\deg(f)$, is the length of the longest word in $\mathrm{supp}(f)$. For a given $d \in \mathbb{N}$, let us denote by $\mathbf{W}_d$ the column vector of all words of degree at most d arranged with respect to. the lexicographic order. The ring $\mathbb{R}\langle \underline{X}\rangle$ is equipped with the involution $\star$ that fixes $\mathbb{R} \cup \{X_1, \ldots, X_n\}$ point-wise and reverses words, so that $\mathbb{R}\langle \underline{X}\rangle$ is the $\star$-algebra freely generated by n symmetric letters $X_1, \ldots, X_n$. The set of symmetric elements in $\mathbb{R}\langle \underline{X}\rangle$ is defined as $\mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle := \{f \in \mathbb{R}\langle \underline{X}\rangle \mid f^\star = f\}$. We use $|\cdot|$ to denote the cardinality of a set and let $[m] := \{1, 2, \ldots, m\}$ for $m \in \mathbb{N}\backslash\{0\}$.

2.2 Sums of hermitian squares

An nc polynomial of the form $g^\star g$ is called a *hermitian square*. An nc polynomial $f \in \mathbb{R}\langle \underline{X}\rangle$ is called a *sum of hermitian squares (SOHS)* if there exist nc polynomials $g_1, \ldots, g_r \in \mathbb{R}\langle \underline{X}\rangle$ such that $f = g_1^\star g_1 + g_2^\star g_2 + \ldots + g_r^\star g_r$. The set of SOHS is denoted by $\Sigma\langle \underline{X}\rangle$. Checking whether a given nc polynomial $f \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ is an SOHS can be cast as a semidefinite program (SDP) due to the following theorem.

Theorem 1 ([19], Theorem 1.1) *Let $f \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ with $\deg(f) = 2d$. Then $f \in \Sigma\langle \underline{X}\rangle$ if and only if there exists a matrix $Q \succeq 0$ satisfying*

$$f = \mathbf{W}_d^\star Q \mathbf{W}_d. \tag{1}$$

Any symmetric matrix Q (not necessarily PSD) satisfying (1) is called a *Gram matrix* of f. The standard monomial basis $\mathbf{W}_d$ used in (1) can be reduced via the Newton chip method; see Chapter 2 in [11].

2.3 Semialgebraic sets and quadratic modules

Given $S = \{g_1, \ldots, g_m\} \subseteq \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$, the *semialgebraic set* $\mathcal{D}_S$ associated with S is defined by

$$\mathcal{D}_S := \bigcup_{r\in\mathbb{N}\setminus\{0\}} \{\underline{A} = (A_1, \ldots, A_n) \in \mathbb{S}_r^n \mid g_j(\underline{A}) \succeq 0, j \in [m]\}. \tag{2}$$

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 $g(\underline{A})$ a PSD operator, for all $g \in S$. The *quadratic module* $\mathcal{M}_S$, generated by S, is defined by

$$\mathcal{M}_S := \{\sum_{j=1}^{s} h_j^\star g_j h_j \mid s \in \mathbb{N}\setminus\{0\}, h_j \in \mathbb{R}\langle \underline{X}\rangle, g_j \in \{1\} \cup S\}, \tag{3}$$

and the *truncated quadratic module* $\mathcal{M}_{S,d}$ of order $d \in \mathbb{N}$, generated by S, is

$$\mathcal{M}_{S,d} := \{\sum_{j=1}^{s} h_j^\star g_j h_j \mid s \in \mathbb{N}\setminus\{0\}, h_j \in \mathbb{R}\langle \underline{X}\rangle, g_j \in \{1\} \cup S, \deg(h_j^\star g_j h_j) \le 2d\}. \tag{4}$$

A quadratic module $\mathcal{M}$ is *Archimedean* if for each $h \in \mathbb{R}\langle \underline{X}\rangle$, there exists $N \in \mathbb{N}$ such that $N - h^\star h \in \mathcal{M}$. The noncommutative analog of Putinar's Positivstellensatz describing noncommutative polynomials positive on $\mathcal{D}_S^\infty$ with Archimedean $\mathcal{M}_S$ is due to Helton and McCullough:

Theorem 2 ([18], Theorem 1.2) *Let* $\{f\} \cup S \subseteq \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ *and assume that* $\mathcal{M}_S$ *is Archimedean. If* $f(A) \succ 0$ *for all* $A \in \mathcal{D}_S^\infty$, *then* $f \in \mathcal{M}_S$.

2.4 Moment and localizing matrices

With $\mathbf{y} = (y_w)_{w\in\langle \underline{X}\rangle}$ being a sequence indexed by the standard monomial basis $\langle \underline{X}\rangle$ of $\mathbb{R}\langle \underline{X}\rangle$, let $L_{\mathbf{y}} : \mathbb{R}\langle \underline{X}\rangle \to \mathbb{R}$ be the linear functional

$$f = \sum_w a_w w \mapsto L_{\mathbf{y}}(f) = \sum_w a_w y_w.$$

Given a monomial basis $\mathbf{B}$, the noncommutative *moment* matrix $M_{\mathbf{B}}(\mathbf{y})$ associated with $\mathbf{B}$ and $\mathbf{y}$ is the matrix with rows and columns indexed by $\mathbf{B}$ such that

$$M_{\mathbf{B}}(\mathbf{y})_{uv} := L_{\mathbf{y}}(u^\star v) = y_{u^\star v}, \quad \forall u, v \in \mathbf{B}.$$

If $\mathbf{B}$ is the standard monomial basis $\mathbf{W}_d$, we also denote $M_{\mathbf{W}_d}(\mathbf{y})$ by $M_d(\mathbf{y})$.

Suppose $g = \sum_w b_w w \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ and let $\mathbf{y} = (y_w)_{w\in\langle \underline{X}\rangle}$ be given. For any positive integer d, the noncommutative *localizing* matrix $M_d(g\mathbf{y})$ associated with g and $\mathbf{y}$ is the matrix with rows and columns indexed by $\mathbf{W}_d$ such that

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

2.5 Eigenvalue optimization for noncommutative polynomials

Given $f = \sum_w a_w w \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$, the eigenvalue minimization problem for f is defined by:

$$\big(\mathrm{EP}_0\big): \quad \lambda_{\min}(f) := \inf\{\langle f\big(\underline{A}\big)\boldsymbol{v}, \boldsymbol{v}\rangle : \underline{A} \in \mathbb{S}_r^n, r \in \mathbb{N}\backslash\{0\}, ||\boldsymbol{v}|| = 1\}. \tag{5}$$

Assume that $\mathbf{B}$ is a monomial basis. Then (EP_0) is equivalent to the following SDP ([11]):

$$(\mathrm{EP}): \quad \begin{aligned} \lambda_{\min}(f) = \ & \inf \ L_{\mathbf{y}}(f) \\ & \text{s.t. } M_{\mathbf{B}}(\mathbf{y}) \succeq 0, \\ & \quad y_1 = 1. \end{aligned} \tag{6}$$

Writing $M_{\mathbf{B}}(\mathbf{y}) = \sum_w A_w y_w$ for appropriate matrices $\{A_w\}_w$, the dual SDP of (6) is

$$\begin{aligned} & \sup \ \lambda \\ & \text{s.t. } \langle Q, A_w\rangle + \lambda\delta_{1w} = a_w, \quad \forall w \in \mathbf{B}^\star\mathbf{B}, \\ & \quad Q \succeq 0, \end{aligned} \tag{7}$$

where $\mathbf{B}^\star\mathbf{B} := \{u^\star v \mid u, v \in \mathbf{B}\}$ and δ_{1w} is the usual Kronecker symbol.

Given $f = \sum_w a_w w \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ and $S = \{g_1, \ldots, g_m\} \subseteq \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$, let us consider the following eigenvalue minimization problem for f over the operator semi-algebraic set $\mathcal{D}_S^\infty$:

$$\big(\mathrm{EQ}_0\big): \quad \lambda_{\min}(f, S) := \inf\{\langle f\big(\underline{A}\big)\boldsymbol{v}, \boldsymbol{v}\rangle : \underline{A} \in \mathcal{D}_S^\infty, ||\boldsymbol{v}|| = 1\}. \tag{8}$$

For convenience, we set $g_0 := 1$ and let $d_j = \lceil \deg(g_j)/2\rceil$ for $j = 0, 1, \ldots, m$. Assume that $\hat{d} \geq d := \max\{\lceil \deg(f)/2\rceil, d_1, \ldots, d_m\}$ is a positive integer. As shown in [34], one has the following hierarchy of moment relaxations, indexed by $\hat{d}$, to obtain a sequence of lower bounds for the optimum $\lambda_{\min}(f, S)$ of (EQ_0):

$$\big(\mathrm{EQ}_{\hat{d}}\big): \quad \begin{aligned} \lambda_{\hat{d}}(f, S) := \ & \inf \ L_{\mathbf{y}}(f) \\ & \text{s.t. } M_{\hat{d}}(\mathbf{y}) \succeq 0, \\ & \quad M_{\hat{d}-d_j}(g_j\mathbf{y}) \succeq 0, \quad j \in [m], \\ & \quad y_1 = 1. \end{aligned} \tag{9}$$

We call $\hat{d}$ the *relaxation order*. If the quadratic module $\mathcal{M}_S$ generated by S is Archimedean, then the sequence of lower bounds $(\lambda_{\hat{d}}(f, S))_{\hat{d}\geq d}$ converges to $\lambda_{\min}(f, S)$. See, e.g., [11, Corollary 4.11] for a proof.

For each j, writing $M_{\hat{d}-d_j}(g_j\mathbf{y}) = \sum_w D_w^j y_w$ for appropriate matrices $\{D_w^j\}_{j,w}$, we can write the dual SDP of (9) as

$$\begin{aligned} &\sup\ \lambda \\ &\text{s.t.}\ \ \sum_{j=0}^{m}\langle Q_j, D_w^j\rangle + \lambda\delta_{1w} = a_w, \quad \forall w \in \mathbf{W}_{2\hat{d}}, \\ &\quad Q_j \succeq 0, \quad j \in \{0\}\cup[m]. \end{aligned} \tag{10}$$

2.6 Trace optimization for noncommutative polynomials

Given $g, h \in \mathbb{R}\langle \underline{X}\rangle$, the nc polynomial $[g,h] := gh - hg$ is called a *commutator*. Two nc polynomials $g, h \in \mathbb{R}\langle \underline{X}\rangle$ are said to be *cyclically equivalent*, denoted by $g \overset{\text{cyc}}{\sim} h$, if $g - h$ is a sum of commutators. Let $w \in \langle \underline{X}\rangle$. The canonical representative $[w]$ of w is the minimal one with respect to the lexicographic order among all words cyclically equivalent to w. For $\mathbf{A} \subseteq \langle \underline{X}\rangle$, $[\mathbf{A}] := \{[w] \mid w \in \mathbf{A}\}$. For an nc polynomial $f = \sum_w a_w w \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle$, the canonical representative of f is defined by $[f] := \sum_w a_w [w] \in \mathbb{R}\langle \underline{X}\rangle$ and the *cyclic degree* of f is defined as $\operatorname{cdeg}(f) := \deg([f])$. We warn the reader about a small abuse of notation as $[k]$ stands for $\{1, \ldots, k\}$ when k is a positive integer.

The *normalized trace* of a matrix $A = [a_{ij}] \in \mathbb{S}_r$ is given by $\operatorname{tr} A = \frac{1}{r}\sum_{i=1}^{r} a_{ii}$. Given $f = \sum_w a_w w \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle$, the trace minimization problem for f is defined by

$$\left(\mathrm{TP}_0\right): \quad \operatorname{tr}_{\min}(f) := \inf\{\operatorname{tr} f(\underline{A}) : \underline{A} \in \mathbb{S}_r^n, r \in \mathbb{N}\backslash\{0\}\}. \tag{11}$$

Let $d = \operatorname{cdeg}(f)$. As shown in [11], (TP_0) admits the following moment relaxation:

$$\begin{aligned} \mu(f) := \ &\inf\ L_{\mathbf{y}}(f) \\ &\text{s.t.}\ M_d(\mathbf{y}) \succeq 0, \\ &\quad M_d(\mathbf{y})_{uv} = M_d(\mathbf{y})_{wz}, \quad \text{for all } u^\star v \overset{\text{cyc}}{\sim} w^\star z, \\ &\quad y_1 = 1. \end{aligned} \tag{12}$$

The dual of (TP) reads as

$$\begin{aligned} &\sup\ \mu \\ &\text{s.t.}\ \ \textstyle\sum_{w \overset{\text{cyc}}{\sim} v}\left(\langle Q, A_w\rangle + \mu\delta_{1w}\right) = \sum_{w \overset{\text{cyc}}{\sim} v} a_w, \quad \forall v \in \mathbf{W}_{2d}, \\ &\quad Q \succeq 0. \end{aligned} \tag{13}$$

Given $f = \sum_w a_w w \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle$ and $S = \{g_1, \ldots, g_m\} \subseteq \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle$, the trace minimization problem for f over the semialgebraic set $\mathcal{D}_S$ is defined by

$$\left(\mathrm{TQ}_0\right): \quad \operatorname{tr}_{\min}(f, S) := \inf\{\operatorname{tr} f(\underline{A}) : \underline{A} \in \mathcal{D}_S\}. \tag{14}$$

We produce lower bounds on $\operatorname{tr}_{\min}(f, S)$ by restricting ourselves to a specific subset of $\mathcal{D}_S^\infty$, obtained by considering the algebra of all bounded operators on a Hilbert space to finite von Neumann algebras [39] of type I and type II. We introduce $\operatorname{tr}_{\min}(f, S)^{\mathrm{II}_1}$ as the trace minimum of f on $\mathcal{D}_S^{\mathrm{II}_1}$. Since $\mathcal{D}_S$ is contained in $\mathcal{D}_S^{\mathrm{II}_1}$, one has $\operatorname{tr}_{\min}(f, S)^{\mathrm{II}_1} \le \operatorname{tr}_{\min}(f, S)$. For a proper definition of $\mathcal{D}_S^{\mathrm{II}_1}$, we refer the interested reader to, e.g., [11, Definition 1.59]. As shown in [10], one has the following series

of moment relaxations indexed by $\hat{d} \geq d$ to obtain a hierarchy of lower bounds for $\mathrm{tr}_{\min}(f,S)^{\mathrm{II}_1}$:

$$\begin{aligned} \mu_{\hat{d}}(f,S) := \ & \inf \ L_{\mathbf{y}}(f) \\ & \text{s.t. } M_{\hat{d}}(\mathbf{y}) \succeq 0, \\ (\mathrm{TQ}_{\hat{d}}) : \quad & M_{\hat{d}-d_j}(g_j\mathbf{y}) \succeq 0, \quad j \in [m], \\ & M_{\hat{d}}(\mathbf{y})_{uv} = M_{\hat{d}}(\mathbf{y})_{wz}, \quad \text{for all } u^\star v \overset{\mathrm{cyc}}{\sim} w^\star z, \\ & y_1 = 1. \end{aligned} \tag{15}$$

We call $\hat{d}$ the *relaxation order*. If the quadratic module $\mathcal{M}_S$ generated by S is Archimedean, then the sequence of lower bounds $(\mu_{\hat{d}}(f,S))_{\hat{d}\geq d}$ converges to $\mathrm{tr}_{\min}(f,S)^{\mathrm{II}_1}$. See, e.g., [11, Corollary 5.5] for a proof.

The dual of (15) reads as

$$\begin{aligned} & \sup \ \mu \\ & \text{s.t.} \ \sum_{w \overset{\mathrm{cyc}}{\sim} v} \left(\sum_{j=0}^{m} \langle Q_j, D_w^j \rangle + \mu \delta_{1w} \right) = \sum_{w \overset{\mathrm{cyc}}{\sim} v} a_w, \quad \forall v \in \mathbf{W}_{2\hat{d}}, \\ & \quad Q_j \succeq 0, \quad j \in \{0\} \cup [m]. \end{aligned} \tag{16}$$

2.7 Chordal graphs and sparse matrices

In this subsection, we briefly revisit the relationship between chordal graphs and sparse matrices, which is crucial for the sparsity-exploitation of this paper. For more details on chordal graphs and sparse matrices, the reader is referred to [40].

An (undirected) *graph* $G(V, E)$ or simply G consists of a set of nodes V and a set of edges $E \subseteq \{\{v_i, v_j\} \mid (v_i, v_j) \in V \times V\}$. When G is a graph, we also use $V(G)$ and $E(G)$ to indicate the node set of G and the edge set of G, respectively. The *adjacency matrix* of G is denoted by B_G for which we put ones on its diagonal. For two graphs G, H, we say that G is a *subgraph* of H if $V(G) \subseteq V(H)$ and $E(G) \subseteq E(H)$, denoted by $G \subseteq H$. For a graph $G(V, E)$, a *cycle* of length k is a set of nodes $\{v_1, v_2, \ldots, v_k\} \subseteq V$ with $\{v_k, v_1\} \in E$ and $\{v_i, v_{i+1}\} \in E$, for $i = 1, \ldots, k-1$. A *chord* in a cycle $\{v_1, v_2, \ldots, v_k\}$ is an edge $\{v_i, v_j\}$ that joins two nonconsecutive nodes in the cycle.

A graph is called a *chordal graph* if all its cycles of length at least four have a chord. Note that any non-chordal graph $G(V, E)$ can always be extended to a chordal graph $\overline{G}(V, \overline{E})$ by adding appropriate edges to E, which is called a *chordal extension* of $G(V, E)$. A *clique* $C \subseteq V$ of G is a subset of nodes where $\{v_i, v_j\} \in E$ for any $v_i, v_j \in C$. If a clique C is not a subset of any other clique, then it is called a *maximal clique*. It is known that maximal cliques of a chordal graph can be enumerated efficiently in linear time in the number of nodes and edges of the graph [5].

Given a graph $G(V, E)$ with $V = [|V|]$, a symmetric matrix Q with rows and columns indexed by V is said to have sparsity pattern G if $Q_{ij} = Q_{ji} = 0$ whenever $i \neq j$ and $\{i, j\} \notin E$, i.e., $B_G \circ Q = Q$. Let $\mathbb{S}_G$ be the set of symmetric matrices with sparsity pattern G. A matrix in $\mathbb{S}_G$ exhibits a *block structure*: each block corresponds to a maximal clique of G. The maximal block size is the maximal size of maximal cliques of

G, namely, the *clique number* of G. Note that there might be overlaps between blocks because different maximal cliques may share nodes.

Given a maximal clique C of $G(V, E)$, we define a matrix $P_C \in \mathbb{R}^{|C|\times|V|}$ as

$$[P_C]_{ij} = \begin{cases} 1, \text{ if } C(i) = j, \\ 0, \text{ otherwise}, \end{cases} \tag{17}$$

where $C(i)$ denotes the i-th node in C, sorted in the ordering compatibly with V. Note that $P_C Q P_C^T \in \mathbb{S}_{|C|}$ extracts a principal submatrix defined by the indices in the clique C from a symmetric matrix $Q \in \mathbb{S}_{|V|}$, and $P_C^T Q_C P_C$ inflates a $|C| \times |C|$ matrix Q_C into a sparse $|V| \times |V|$ matrix.

The PSD matrices with sparsity pattern G form a convex cone

$$\mathbb{S}^+_{|V|} \cap \mathbb{S}_G = \{Q \in \mathbb{S}_G \mid Q \succeq 0\}. \tag{18}$$

When the sparsity pattern graph G is chordal, the cone $\mathbb{S}^+_{|V|} \cap \mathbb{S}_G$ can be decomposed as a sum of simple convex cones, as stated in the following theorem.

Theorem 3 ([1], see also [40]) *Let $G(V, E)$ be a chordal graph and assume that $C_1, \ldots, C_t$ are the list of maximal cliques of $G(V, E)$. Then a matrix $Q \in \mathbb{S}^+_{|V|} \cap \mathbb{S}_G$ if and only if there exists $Q_i \in \mathbb{S}^+_{|C_i|}$ for $i = 1, \ldots, t$ such that $Q = \sum_{i=1}^t P_{C_i}^T Q_i P_{C_i}$.*

Given a graph $G(V, E)$ with $V = [|V|]$, let Π_G be the projection from $\mathbb{S}_{|V|}$ to the subspace $\mathbb{S}_G$, i.e., for $Q \in \mathbb{S}_{|V|}$,

$$\Pi_G(Q)_{ij} = \begin{cases} Q_{ij}, \text{ if } \{i, j\} \in E \text{ or } i = j, \\ 0, \quad \text{otherwise}. \end{cases} \tag{19}$$

We denote by $\Pi_G(\mathbb{S}^+_{|V|})$ the set of matrices in $\mathbb{S}_G$ that have a PSD completion, i.e.,

$$\Pi_G\left(\mathbb{S}^+_{|V|}\right) = \{\Pi_G(Q) \mid Q \in \mathbb{S}^+_{|V|}\}. \tag{20}$$

One can check that the PSD completable cone $\Pi_G(\mathbb{S}^+_{|V|})$ and the PSD cone $\mathbb{S}^+_{|V|} \cap \mathbb{S}_G$ form a pair of dual cones in $\mathbb{S}_G$; see [40, Section 10.1] for a proof. Moreover, for a chordal graph G, the decomposition result for the cone $\mathbb{S}^+_{|V|} \cap \mathbb{S}_G$ in Theorem 3 leads to the following characterization of the PSD completable cone $\Pi_G(\mathbb{S}^+_{|V|})$.

Theorem 4 ([17], see also [40]) *Let $G(V, E)$ be a chordal graph and assume that $C_1, \ldots, C_t$ are the list of maximal cliques of $G(V, E)$. Then a matrix $Q \in \Pi_G(\mathbb{S}^+_{|V|})$ if and only if $Q_i = P_{C_i} Q P_{C_i}^T \succeq 0$ for $i = 1, \ldots, t$.*

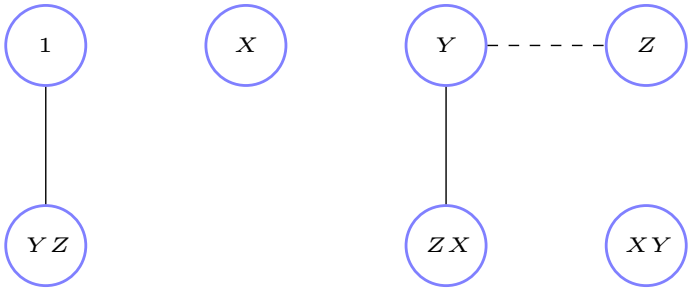

Fig. 1 The support extension SE(G) of G for Example 1. The dashed edges are added after the support extension

3 Eigenvalue optimization for noncommutative polynomials with term sparsity

In this section, we consider the eigenvalue optimization problem for noncommutative polynomials with term sparsity. For the reader's convenience, we first deal with the unconstrained case and then generalize to the constrained case.

3.1 The unconstrained case

In this subsection, we describe an iterative procedure to exploit term sparsity for the moment-SOHS relaxations (6–7) of the unconstrained NCPOP (EP_0) defined in (5).

Let $f = \sum_{w\in\mathbf{A}} a_w w \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ with $\mathrm{supp}(f) = \mathbf{A}$ (w.l.o.g. assuming $1 \in \mathbf{A}$). Assume that $\mathbf{B}$ is the monomial basis returned by the Newton chip method ([11, §2.3]) with $r = |\mathbf{B}|$. To represent the term sparsity in f, in the sequel we will consider graphs with $V := \mathbf{B}$ as the set of nodes. Suppose that $G(V, E)$ is such a graph. We define the *support* of G by

$$\mathrm{supp}(G) := \{u^\star v \mid (u, v) \in V \times V,\ \{u, v\} \in E\}.$$

We further define two operations on G: *support extension* and *chordal extension*.

(1) *support extension* The support extension of G, denoted by SE(G), is the graph with nodes $\mathbf{B}$ and with edges

$$E(\mathrm{SE}(G)) := \{\{u, v\} \mid (u, v) \in V \times V,\ u \neq v,\ u^\star v \in \mathrm{supp}(G) \cup \mathbf{B}^2\},$$

where $\mathbf{B}^2 := \{u^\star u \mid u \in \mathbf{B}\}$.

Example 1 Consider the graph $G(V, E)$ with

$$V = \{1, X, Y, Z, YZ, ZX, XY\} \text{ and } E = \{\{1, YZ\}, \{Y, ZX\}\}.$$

Then $E(\mathrm{SE}(G)) = \{\{1, YZ\}, \{Y, ZX\}, \{Y, Z\}\}$. See Fig. 1 for the support extension SE(G) of G.

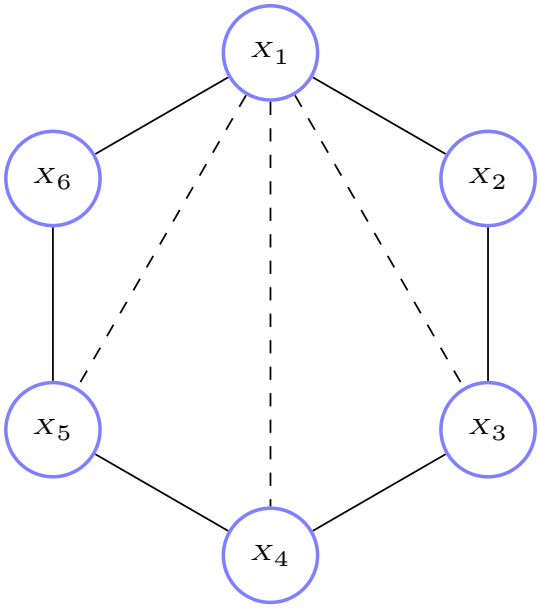

Fig. 2 A smallest chordal extension $\overline{G}$ of G for Example 2. The dashed edges are added after the chordal extension

(2) *Chordal extension*: For a graph G, we denote any specific chordal extension of G by $\overline{G}$. There are generally various chordal extensions of G. In this paper, we will consider two particular types of chordal extensions: *the maximal chordal extension* and *approximately smallest chordal extensions*. By the maximal chordal extension, we refer to the chordal extension that completes every connected component of G. The maximal chordal extension can be easily computed by listing all connected components. Another advantage of the maximal chordal extension is that there is no overlap among maximal cliques. However, the clique number of the maximal chordal extension may be large among all possible chordal extensions. A chordal extension with the smallest clique number is called a *smallest chordal extension*. Computing a smallest chordal extension of a graph is an NP-complete problem in general. Fortunately, several heuristic algorithms, e.g., the greedy minimum degree and the greedy minimum fill-ins, are known to efficiently produce a good approximation; see [6] for more detailed discussions. Throughout the paper, we assume that for graphs G, H,

$$G \subseteq H \Longrightarrow \overline{G} \subseteq \overline{H}. \tag{21}$$

This assumption is reasonable since any chordal extension of H restricting to G is also a chordal extension of G.

Example 2 Consider the graph $G(V, E)$ with $V = \{X_1, X_2, X_3, X_4, X_5, X_6\}$ and $E = \{\{X_1, X_2\}, \{X_2, X_3\}, \{X_3, X_4\}, \{X_4, X_5\}, \{X_5, X_6\}, \{X_6, X_1\}\}$. See Fig. 2 for a smallest chordal extension $\overline{G}$ of G which has 4 maximal cliques of size 3. On the other hand, the maximal chordal extension of G has one maximal clique of size 6.

Now we define $G_0(V, E_0)$ to be the graph with $V = \mathbf{B}$ and

$$E_0 = \big\{\{u, v\} \mid (u, v) \in V \times V,\ u \neq v,\ u^\star v \in \mathbf{A} \cup \mathbf{B}^2\big\}, \tag{22}$$

which is called the *term sparsity pattern (tsp) graph* associated with f. We then iteratively define a sequence of graphs $(G_k(V, E_k))_{k\geq 1}$ by alternately performing support extension and chordal extension to $G_0(V, E_0)$:

$$G_k := \overline{\mathrm{SE}(G_{k-1})}. \tag{23}$$

When f is sparse (i.e., G_1 is not complete), by replacing $M_{\mathbf{B}}(\mathbf{y}) \succeq 0$ with the weaker condition $B_{G_k} \circ M_{\mathbf{B}}(\mathbf{y}) \in \Pi_{G_k}(\mathbb{S}_r^+)$ in (6), we obtain a series of sparse moment relaxations of (EP) (and (EP_0)) indexed by $k \geq 1$:

$$\begin{aligned} \lambda_k(f) := \ & \inf\ L_{\mathbf{y}}(f) \\ (\mathrm{EP}^k)\ : \quad & \text{s.t. } B_{G_k} \circ M_{\mathbf{B}}(\mathbf{y}) \in \Pi_{G_k}(\mathbb{S}_r^+), \\ & \quad y_1 = 1. \end{aligned} \tag{24}$$

We call k the *sparse order*. By construction, one has $G_k \subseteq G_{k+1}$ for all $k \geq 1$ and therefore the sequence of graphs $(G_k(V, E_k))_{k\geq 1}$ stabilizes after a finite number of steps. We denote the stabilized graph by $G_\bullet(V, E_\bullet)$ and the corresponding moment relaxation by $(\mathrm{EP}^\bullet)$ (with optimum $\lambda_\bullet(f)$).

For each $k \geq 1$, the dual SDP of (24) reads as

$$\begin{aligned} \sup\ & \lambda \\ \text{s.t. } & \langle Q, A_w \rangle + \lambda \delta_{1w} = a_w, \quad \forall w \in \mathrm{supp}(G_k) \cup \mathbf{B}^2, \\ & Q \in \mathbb{S}_r^+ \cap \mathbb{S}_{G_k}, \end{aligned} \tag{25}$$

where A_w is defined in Sect. 2.5.

Theorem 5 *Assume that $f \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X} \rangle$. Then the following hold*:

(1) *For each $k \geq 1$, there is no duality gap between* (EP^k) (24) *and its dual* (25).
(2) *The sequence $(\lambda_k(f))_{k\geq 1}$ is monotone nondecreasing and $\lambda_k(f) \leq \lambda_{\min}(f)$ for all k (with $\lambda_{\min}(f)$ defined in* (6)).
(3) *If the maximal chordal extension is used in* (23), *then $(\lambda_k(f))_{k\geq 1}$ converges to $\lambda_{\min}(f)$ in finitely many steps, i.e., $\lambda_\bullet(f) = \lambda_{\min}(f)$.*

Proof (i) Note that the SDP problem (EP) (6) has a Slater's point, i.e., a strictly feasible solution (see, e.g., Proposition 4.9 in [11]), say $M_{\mathbf{B}}(\mathbf{y}^{\mathrm{opt}})$. Since each block of $B_{G_k} \circ M_{\mathbf{B}}(\mathbf{y}^{\mathrm{opt}})$ is a principal submatrix of $M_{\mathbf{B}}(\mathbf{y}^{\mathrm{opt}})$, we have that $B_{G_k} \circ M_{\mathbf{B}}(\mathbf{y}^{\mathrm{opt}})$ is a Slater's point of (EP^k) by Theorem 4. So by the duality theory of convex programming, there is no duality gap between (EP^k) and its dual.

(ii) Because $G_k \subseteq G_{k+1}$, each maximal clique of G_k is a subset of some maximal clique of G_{k+1}. Thus by Theorem 4, we have that (EP^k) is a relaxation of (EP^{k+1}) (and also a relaxation of (EP)). This yields the desired conclusions.

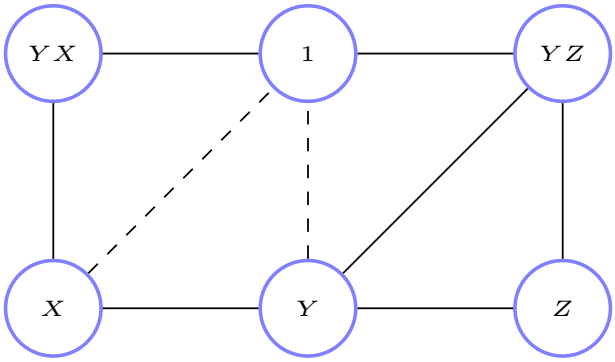

Fig. 3 The tsp graph G_0 and its chordal extension G_1 for Example 3

(iii) Let $\mathbf{y}^{\text{opt}} = (y_w^{\text{opt}})_w$ be an arbitrary feasible solution of (EP$^\bullet$). Note that $\{y_w \mid w \in \text{supp}(G_\bullet) \cup \mathbf{B}^2\}$ is the set of decision variables involved in (EP$^\bullet$) and $\{y_w \mid w \in \mathbf{B}^\star\mathbf{B}\}$ is the set of decision variables involved in (EP). We then define a vector $\overline{\mathbf{y}}^{\text{opt}} = (\overline{y}_w^{\text{opt}})_{w\in\mathbf{B}^\star\mathbf{B}}$ as follows:

$$\overline{y}_w^{\text{opt}} = \begin{cases} y_w^{\text{opt}}, & \text{if } w \in \text{supp}(G_\bullet) \cup \mathbf{B}^2, \\ 0, & \text{otherwise.} \end{cases}$$

If the maximal chordal extension is used in (23), then matrices in $\Pi_{G_k}(\mathbb{S}_r^+)$ for all $k \geq 1$ are block-diagonal (up to permutation). As a consequence, $B_{G_k} \circ M_{\mathbf{B}}(\mathbf{y}) \in \Pi_{G_k}(\mathbb{S}_r^+)$ implies $B_{G_k} \circ M_{\mathbf{B}}(\mathbf{y}) \succeq 0$. By construction, we have $M_{\mathbf{B}}(\overline{\mathbf{y}}^{\text{opt}}) = B_{G_\bullet} \circ M_{\mathbf{B}}(\mathbf{y}^{\text{opt}}) \succeq 0$. Therefore $\overline{\mathbf{y}}^{\text{opt}}$ is a feasible solution of (EP) and hence $L_{\mathbf{y}^{\text{opt}}}(f) = L_{\overline{\mathbf{y}}^{\text{opt}}}(f) \geq \lambda_{\min}(f)$. This yields $\lambda_\bullet(f) \geq \lambda_{\min}(f)$ since $\mathbf{y}^{\text{opt}}$ is an arbitrary feasible solution of (EP$^\bullet$). By (ii), we already have $\lambda_\bullet(f) \leq \lambda_{\min}(f)$. Therefore, $\lambda_\bullet(f) = \lambda_{\min}(f)$. □

If approximately smallest chordal extensions are used in (23), the sequence $(\lambda_k(f))_{k\geq 1}$ doesn't necessarily converge to $\lambda_{\min}(f)$. The following is an example.

Example 3 Consider the nc polynomial $f = X^2 - XY - YX + 3Y^2 - 2XYX + 2XY^2X - YZ - ZY + 6Z^2 + 9X^2Y + 9Z^2Y - 54ZYZ + 142ZY^2Z$ ([21]). The monomial basis given by the Newton chip method is $\{1, X, Y, Z, YX, YZ\}$. We have $E_0 = \{\{1, YX\}, \{1, YZ\}, \{X, YX\}, \{X, Y\}, \{Y, Z\}, \{Y, YZ\}, \{Z, YZ\}\}$. Figure 3 shows the tsp graph G_0 (without dashed edges) and its chordal extension G_1 (with dashed edges) for f. The graph sequence $(G_k)_{k\geq 1}$ immediately stabilizes at $k = 1$. Solving the SDP problem (EP^1) associated with G_1, we obtain $\lambda_1(f) \approx -0.00355$ whereas we have $\lambda_{\min}(f) = 0$.

The next result states that $\lambda_1(f) = \lambda_{\min}(f)$ always holds for a quadratic f.

Proposition 1 *Suppose that the nc polynomial $f \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ in* (EP_0) *is quadratic, i.e.*, $\deg(f) = 2$. *Then* $\lambda_1(f) = \lambda_{\min}(f)$.

Proof Let $\mathbf{A} = \mathrm{supp}(f)$. As f is quadratic, we may take $\mathbf{B} = \{1, X_1, \ldots, X_n\}$ as a monomial basis. Let G_0 be the tsp graph associated with f. We only need to prove that if f admits a PSD Gram matrix, then f admits a Gram matrix in $\mathbb{S}^+_{n+1} \cap \mathbb{S}_{G_0}$. Suppose that $Q = [q_{ij}]^n_{i,j=0}$ is a PSD Gram matrix for f indexed by $\mathbf{B}$. Note that for $i, j > 0$, if $\{X_i, X_j\} \notin E(G_0)$, then we must have $X_iX_j, X_jX_i \notin \mathbf{A}$, which implies $q_{ij} = 0$; for $i = 0, j > 0$, if $\{1, X_j\} \notin E(G_0)$, then we must have $X_j \notin \mathbf{A}$, which implies $q_{0j} = q_{j0} = 0$. It follows that $Q \in \mathbb{S}_{G_0}$ as desired. □

3.2 The constrained case

In this subsection, we generalize the iterative procedure in Sect. 3.1 to the constrained case and we show how to iteratively exploit term sparsity for the moment-SOHS hierarchy (9–10) of the constrained NCPOP (EQ_0) defined in (8).

Assume that $f = \sum_w a_w w \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ and $S = \{g_1, \ldots, g_m\} \subseteq \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$. Let

$$\mathbf{A} = \mathrm{supp}(f) \cup \bigcup_{j=1}^{m} \mathrm{supp}(g_j). \tag{26}$$

As in Sect. 2.5, we set $g_0 := 1$, and let $d_j = \lceil \deg(g_j)/2 \rceil$, $j \in \{0\} \cup [m]$ and $d = \max\{\lceil \deg(f)/2 \rceil, d_1, \ldots, d_m\}$. Fixing a relaxation order $\hat{d} \geq d$, we define a graph $G^{\mathrm{tsp}}_{\hat{d}}(V^{\mathrm{tsp}}_{\hat{d}}, E^{\mathrm{tsp}}_{\hat{d}})$ with $V^{\mathrm{tsp}}_{\hat{d}} = \mathbf{W}_{\hat{d}}$ and

$$E^{\mathrm{tsp}}_{\hat{d}} = \{\{u, v\} \mid (u, v) \in \mathbf{W}_{\hat{d}} \times \mathbf{W}_{\hat{d}},\ u \neq v,\ u^\star v \in \mathbf{A} \cup \mathbf{W}^2_{\hat{d}}\}, \tag{27}$$

where $\mathbf{W}^2_{\hat{d}} := \{u^\star u \mid u \in \mathbf{W}_{\hat{d}}\}$. We call $G^{\mathrm{tsp}}_{\hat{d}}$ the *term sparsity pattern (tsp) graph* associated with $\mathbf{A}$ (or f and S).

For a graph $G(V, E)$ with $V \subseteq \langle \underline{X}\rangle$ and $g \in \mathbb{R}\langle \underline{X}\rangle$, let us define

$$\mathrm{supp}_g(G) := \{u^\star w v \mid (u, v) \in V \times V,\ \{u, v\} \in E, w \in \mathrm{supp}(g)\}. \tag{28}$$

Let $G^{(0)}_{\hat{d},0} = G^{\mathrm{tsp}}_{\hat{d}}$ and $G^{(0)}_{\hat{d},j}$ be the empty graph for $j \in [m]$. Then for each $j \in \{0\} \cup [m]$, we iteratively define a sequence of graphs $(G^{(k)}_{\hat{d},j}(V_{\hat{d},j}, E^{(k)}_{\hat{d},j}))_{k\geq 1}$ with $V_{\hat{d},j} = \mathbf{W}_{\hat{d}-d_j}$ via two successive operations:

(1) *support extension* Let $F^{(k)}_{\hat{d},j}$ be the graph with $V(F^{(k)}_{\hat{d},j}) = \mathbf{W}_{\hat{d}-d_j}$ and

$$\begin{aligned} E\left(F^{(k)}_{\hat{d},j}\right) = \{\{u, v\} \mid (u, v) \in \mathbf{W}_{\hat{d}-d_j} \times \mathbf{W}_{\hat{d}-d_j},\ u \neq v,\\ u^\star \mathrm{supp}(g_j) v \cap \left(\bigcup_{j=0}^{m} \mathrm{supp}_{g_j}\left(G^{(k-1)}_{\hat{d},j}\right) \cup \mathbf{W}^2_{\hat{d}}\right) \neq \emptyset\}, \end{aligned} \tag{29}$$

where $u^\star \mathrm{supp}(g_j) v := \{u^\star w v \mid w \in \mathrm{supp}(g_j)\}$.

(2) *chordal extension*: Let

$$G^{(k)}_{\hat{d},j} := \overline{F^{(k)}_{\hat{d},j}}. \tag{30}$$

Let $r_j = |\mathbf{W}_{\hat{d}-d_j}|$ for $j \in \{0\} \cup [m]$. Then by replacing $M_{\hat{d}-d_j}(g_j\mathbf{y}) \succeq 0$ with the weaker condition $B_{G^{(k)}_{\hat{d},j}} \circ M_{\hat{d}-d_j}(g_j\mathbf{y}) \in \Pi_{G^{(k)}_{\hat{d},j}}(\mathbb{S}^+_{r_j})$ for $j \in \{0\} \cup [m]$ in (9), we obtain the following series of sparse moment relaxations for $(\mathrm{EQ}_{\hat{d}})$ indexed by $k \geq 1$:

$$\left(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},k}\right): \begin{array}{rl} \lambda^{\mathrm{ts}}_{\hat{d},k}(f,S) := & \inf\ L_{\mathbf{y}}(f) \\ & \text{s.t. } B_{G^{(k)}_{\hat{d},0}} \circ M_{\hat{d}}(\mathbf{y}) \in \Pi_{G^{(k)}_{\hat{d},0}}\left(\mathbb{S}^+_{r_0}\right), \\ & \quad B_{G^{(k)}_{\hat{d},j}} \circ M_{\hat{d}-d_j}(g_j\mathbf{y}) \in \Pi_{G^{(k)}_{\hat{d},j}}\left(\mathbb{S}^+_{r_j}\right), \quad j \in [m], \\ & \quad y_1 = 1. \end{array} \tag{31}$$

We call k the *sparse order*. By construction, one has $G^{(k)}_{\hat{d},j} \subseteq G^{(k+1)}_{\hat{d},j}$ for all j, k. Therefore, for every j, the sequence of graphs $(G^{(k)}_{\hat{d},j})_{k\geq 1}$ stabilizes after a finite number of steps. We denote the stabilized graphs by $G^{(\bullet)}_{\hat{d},j}$ for all j and denote the corresponding moment relaxation by $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},\bullet})$ (with optimum $\lambda^{\mathrm{ts}}_{\hat{d},\bullet}(f,S)$).

For each $k \geq 1$, the dual of $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},k})$ reads as

$$\begin{array}{ll} \sup & \lambda \\ \text{s.t.} & \sum_{j=0}^m \langle Q_j, D^j_w \rangle + \lambda\delta_{1w} = a_w, \quad \forall w \in \bigcup_{j=0}^m \mathrm{supp}_{g_j}(G^{(k)}_{\hat{d},j})) \cup \mathbf{W}^2_{\hat{d}}, \\ & Q_j \in \mathbb{S}^+_{r_j} \cap \mathbb{S}_{G^{(k)}_{\hat{d},j}}, \quad j \in \{0\} \cup [m], \end{array} \tag{32}$$

where $\{D^j_w\}_{j,w}$ is defined in Sect. 2.5.

For $\varepsilon > 0$, we define the *nc ε-neighborhood* of 0 to be

$$\mathcal{N}_\varepsilon := \bigcup_{r\in\mathbb{N}\setminus\{0\}} \{\underline{A} = (A_1, \ldots, A_n) \in \mathbb{S}^n_r \mid \varepsilon^2 - \sum_{i=1}^n A_i^2 \succeq 0\}. \tag{33}$$

Theorem 6 *Let $\{f\} \cup S \subseteq \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$. Then the following hold*:

(1) *Suppose that $\mathcal{D}_S$ contains an nc ε-neighborhood of 0.*
Then for all $\hat{d}, k$, there is no duality gap between $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},k})$ (31) and its dual (32).

(2) *Fixing a relaxation order $\hat{d} \geq d$, the sequence $(\lambda^{\mathrm{ts}}_{\hat{d},k}(f,S))_{k\geq 1}$ is monotone nondecreasing and $\lambda^{\mathrm{ts}}_{\hat{d},k}(f,S) \leq \lambda_{\hat{d}}(f,S)$ for all k (with $\lambda_{\hat{d}}(f,S)$ defined in (9)).*

(3) *Fixing a sparse order $k \geq 1$, the sequence $(\lambda^{\mathrm{ts}}_{\hat{d},k}(f,S))_{\hat{d}\geq d}$ is monotone nondecreasing.*

(4) *If the maximal chordal extension is used in (30), then $(\lambda^{\mathrm{ts}}_{\hat{d},k}(f,S))_{k\geq 1}$ converges to $\lambda_{\hat{d}}(f,S)$ in finitely many steps, i.e., $\lambda^{\mathrm{ts}}_{\hat{d},\bullet}(f,S) = \lambda_{\hat{d}}(f,S)$.*

Proof (i) By Proposition 4.9 in [11], the SDP $(\mathrm{EQ}_{\hat{d}})$ (9) admits a Slater's point, say $\mathbf{y}^{\mathrm{opt}}$. Then $M_{\hat{d}}(\mathbf{y}^{\mathrm{opt}}), M_{\hat{d}-d_j}(g_j\mathbf{y}^{\mathrm{opt}})$ are positive definite. Since each block of $B_{G^{(k)}_{\hat{d},0}} \circ M_{\hat{d}}(\mathbf{y}^{\mathrm{opt}})$ (resp. $B_{G^{(k)}_{\hat{d},j}} \circ M_{\hat{d}-d_j}(g_j\mathbf{y}^{\mathrm{opt}})$) is a principal submatrix of $M_{\hat{d}}(\mathbf{y}^{\mathrm{opt}})$ (resp. $M_{\hat{d}-d_j}(g_j\mathbf{y}^{\mathrm{opt}})$), we have that $B_{G^{(k)}_{\hat{d},0}} \circ M_{\hat{d}}(\mathbf{y}^{\mathrm{opt}}), B_{G^{(k)}_{\hat{d},j}} \circ M_{\hat{d}-d_j}(g_j\mathbf{y}^{\mathrm{opt}})$ correspond to a Slater's point of $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},k})$ by Theorem 4. So by the duality theory of convex programming, there is no duality gap between $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},k})$ and its dual.

(ii) For all j, k, because $G^{(k)}_{\hat{d},j} \subseteq G^{(k+1)}_{\hat{d},j}$, each maximal clique of $G^{(k)}_{\hat{d},j}$ is a subset of some maximal clique of $G^{(k+1)}_{\hat{d},j}$. Hence by Theorem 4, $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},k})$ is a relaxation of $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},k+1})$ (and also a relaxation of $(\mathrm{EQ}_{\hat{d}})$). Therefore, $(\lambda^{\mathrm{ts}}_{\hat{d},k}(f,S))_{k\geq 1}$ is monotone non-decreasing and $\lambda^{\mathrm{ts}}_{\hat{d},k}(f,S) \leq \lambda_{\hat{d}}(f,S)$ for all k.

(iii) The conclusion follows if we can show that $G^{(k)}_{\hat{d},j} \subseteq G^{(k)}_{\hat{d}+1,j}$ for all $\hat{d}, j$ since by Theorem 4 this implies that $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},k})$ is a relaxation of $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d}+1,k})$. Let us prove $G^{(k)}_{\hat{d},j} \subseteq G^{(k)}_{\hat{d}+1,j}$ by induction on k. For $k = 1$, from (27), we have $E^{(0)}_{\hat{d},0} \subseteq E^{(0)}_{\hat{d}+1,0}$, which implies that $G^{(1)}_{\hat{d},j} \subseteq G^{(1)}_{\hat{d}+1,j}$ for all $\hat{d}, j$. Now assume that $G^{(k)}_{\hat{d},j} \subseteq G^{(k)}_{\hat{d}+1,j}$ for all $\hat{d}, j$ hold for a given $k \geq 1$. Then from (21), (29), (30) and by the induction hypothesis, we have $G^{(k+1)}_{\hat{d},j} \subseteq G^{(k+1)}_{\hat{d}+1,j}$ for all $\hat{d}, j$, which completes the induction and also completes the proof.

(iv) Let $\mathbf{y}^{\mathrm{opt}} = (y^{\mathrm{opt}}_w)_w$ be an arbitrary feasible solution of $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},\bullet})$. Note that $\{y_w \mid w \in \bigcup_{i=0}^m \mathrm{supp}_{g_i}(G^{(\bullet)}_{\hat{d},j}) \cup \mathbf{W}^2_{\hat{d}}\}$ is the set of decision variables involved in $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},\bullet})$ and $\{y_w \mid w \in \mathbf{W}^{\star}_{\hat{d}}\mathbf{W}_{\hat{d}}\}$ is the set of decision variables involved in $(\mathrm{EQ}_{\hat{d}})$. We then define a vector $\overline{\mathbf{y}}^{\mathrm{opt}} = (\overline{y}^{\mathrm{opt}}_w)_{w\in\mathbf{W}^{\star}_{\hat{d}}\mathbf{W}_{\hat{d}}}$ as follows:

$$\overline{y}^{\mathrm{opt}}_w = \begin{cases} y^{\mathrm{opt}}_w, & \text{if } w \in \bigcup_{i=0}^m \mathrm{supp}_{g_i}(G^{(\bullet)}_{\hat{d},j}) \cup \mathbf{W}^2_{\hat{d}}, \\ 0, & \text{otherwise.} \end{cases}$$

If the maximal chordal extension is used in (30), then the matrices in $\Pi_{G^{(k)}_{\hat{d},j}}(\mathbb{S}^+_{r_j})$ for all $k \geq 1$ are block-diagonal (up to permutation). As a consequence, $B_{G^{(k)}_{\hat{d},j}} \circ M_{\hat{d}-d_j}(g_j\mathbf{y}) \in \Pi_{G^{(k)}_{\hat{d},j}}(\mathbb{S}^+_{r_j})$ implies $B_{G^{(k)}_{\hat{d},j}} \circ M_{\hat{d}-d_j}(g_j\mathbf{y}) \succeq 0$. By construction, we have $M_{\hat{d}-d_j}(g_j\overline{\mathbf{y}}^{\mathrm{opt}}) = B_{G^{(\bullet)}_{\hat{d},j}} \circ M_{\hat{d}-d_j}(g_j\mathbf{y}^{\mathrm{opt}}) \succeq 0$ for all $j \in \{0\} \cup [m]$. Therefore $\overline{\mathbf{y}}^{\mathrm{opt}}$ is a feasible solution of $(\mathrm{EQ}_{\hat{d}})$ and hence $L_{\mathbf{y}^{\mathrm{opt}}}(f) = L_{\overline{\mathbf{y}}^{\mathrm{opt}}}(f) \geq \lambda_{\hat{d}}(f,S)$, which yields $\lambda^{\mathrm{ts}}_{\hat{d},\bullet}(f,S) \geq \lambda_{\hat{d}}(f,S)$ since $\mathbf{y}^{\mathrm{opt}}$ is an arbitrary feasible solution of $(\mathrm{EQ}^{\mathrm{ts}}_{\hat{d},\bullet})$. By (ii), we already have $\lambda^{\mathrm{ts}}_{\hat{d},\bullet}(f,S) \leq \lambda_{\hat{d}}(f,S)$. Therefore, $\lambda^{\mathrm{ts}}_{\hat{d},\bullet}(f,S) = \lambda_{\hat{d}}(f,S)$. □

Following from Theorem 6, we have the following two-level hierarchy of lower bounds for the optimum $\lambda_{\min}(f,S)$ of (EQ_0):

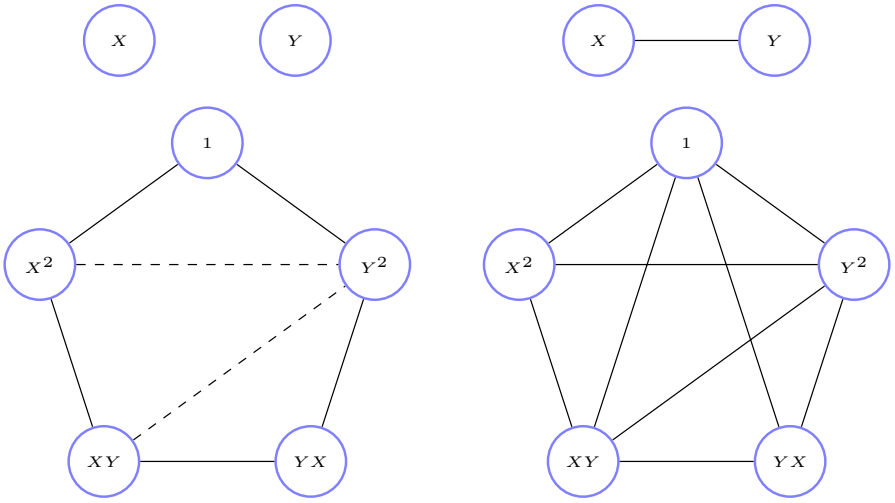

Fig. 4 The graph sequence $(G_{2,0}^{(k)})_{k\geq 1}$ in Example 4: left for $k = 1$; right for $k = 2$. The dashed edges are added after a chordal extension

$$\begin{array}{ccccccc}
\lambda_{d,1}^{\mathrm{ts}}(f,S) & \leq & \lambda_{d,2}^{\mathrm{ts}}(f,S) & \leq \cdots \leq & \lambda_{d}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\lambda_{d+1,1}^{\mathrm{ts}}(f,S) & \leq & \lambda_{d+1,2}^{\mathrm{ts}}(f,S) & \leq \cdots \leq & \lambda_{d+1}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\vdots & & \vdots & \vdots & \vdots \\
\wedge| & & \wedge| & & \wedge| \\
\lambda_{\hat{d},1}^{\mathrm{ts}}(f,S) & \leq & \lambda_{\hat{d},2}^{\mathrm{ts}}(f,S) & \leq \cdots \leq & \lambda_{\hat{d}}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\vdots & & \vdots & \vdots & \vdots
\end{array} \tag{34}$$

We call the array of lower bounds (34) (and its corresponding moment-SOHS relaxations (31–32)) the NCTSSOS hierarchy associated with (EQ_0).

Remark 1 The NCTSSOS hierarchy entails a trade-off between the computational cost and the quality of the obtained lower bound via the two parameters $\hat{d}$ and k. Besides, one has the freedom to choose a specific chordal extension for any graph involved in (30) (e.g., maximal chordal extensions, approximately smallest chordal extensions and so on). This choice affects the resulting sizes of SDP blocks and the quality of the lower bound given by the corresponding SDP relaxation. Intuitively, chordal extensions with smaller clique numbers should lead to SDP blocks of smaller sizes and lower bounds of (possibly) lower quality while chordal extensions with larger clique numbers should lead to SDP blocks with larger sizes and lower bounds of (possibly) higher quality.

Example 4 Consider $f = 2 - X^2 + XY^2X - Y^2 + XYXY + YXYX + X^3Y + YX^3 + XY^3 + Y^3X$ and $S = \{1 - X^2, 1 - Y^2\}$. The graph sequence $(G^{(k)}_{2,0})_{k\geq 1}$ for f and S is given in Fig. 4. In fact the graph sequence $(G^{(k)}_{2,j})_{k\geq 1}$ stabilizes at $k = 2$ for all j (with approximately smallest chordal extensions). Using `NCTSSOS`, we obtain that $\lambda^{\mathrm{ts}}_{2,1}(f, S) \approx -2.55482$, $\lambda^{\mathrm{ts}}_{2,2}(f, S) = \lambda_2(f, S) \approx -2.05111$.

Remark 2 In the case of eigenvalue minimization over nc balls/multi-balls (as in the above example), the first dense relaxation is exact; see Proposition 4.16 and Proposition 4.17 from [11]. We believe that there is no analog of this finite convergence result for term sparsity, i.e., the first or the second step is not always exact. Actually, we have tested this on randomly generated examples of degree 4 (as we do later on for Table 9 in Sect. 6 but without considering correlative sparsity). When using the minimum relaxation order 2 and maximal chordal extensions, the NCTSSOS hierarchy takes a few (typically $2 \sim 3$) steps to converge to the value of the dense relaxation (i.e., the exact value in this case). The sizes of the resulting SDPs heavily depend on the considered problem.

4 Eigenvalue optimization for noncommutative polynomials with combined correlative-term sparsity

The exploitation of term sparsity developed in the previous section can be combined with the exploitation of correlative sparsity discussed in [21] to reduce the computational cost further. To begin with, let us recall some basics on correlative sparsity. For more details, the reader is referred to [21].

4.1 Eigenvalue optimization for noncommutative polynomials with correlative sparsity

As in the commutative case, the exploitation of correlative sparsity in the moment-SOHS hierarchy for NCPOPs consists of two steps: (1) partition the set of variables into subsets according to the correlations between variables emerging in the problem, and (2) construct a sparse moment-SOHS hierarchy with respect to the former partition of variables [21].

More concretely, assuming $f = \sum_w a_w w \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ and $S = \{g_1, \ldots, g_m\} \subseteq \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$, we define the *correlative sparsity pattern (csp) graph* associated with f and S to be the graph G^{csp} with nodes $V = [n]$ and edges E satisfying $\{i, j\} \in E$ if one of following holds:

(1) there exists $w \in \mathrm{supp}(f)$ s.t. $X_i, X_j \in \mathrm{var}(w)$;
(2) there exists l, with $1 \leq l \leq m$, s.t. $X_i, X_j \in \mathrm{var}(g_l)$,

where we use $\mathrm{var}(g)$ to denote the set of variables effectively involved in $g \in \mathbb{R}\langle \underline{X}\rangle$. Let $\overline{G}^{\mathrm{csp}}$ be a chordal extension of G^{csp} and $I_l, l \in [p]$ be the maximal cliques of $\overline{G}^{\mathrm{csp}}$ with cardinality denoted by $n_l, l \in [p]$. Let $\mathbb{R}\langle \underline{X}(I_l)\rangle$ denote the ring of nc polynomials in the n_l variables $\underline{X}(I_l) = \{X_i \mid i \in I_l\}$. We then partition the constraints $g_1, \ldots, g_m$ into groups $\{g_j \mid j \in J_l\}, l \in [p]$ which satisfy

(1) $J_1, \ldots, J_p \subseteq [m]$ are pairwise disjoint and $\bigcup_{l=1}^{p} J_l = [m]$;
(2) for any $j \in J_l$, $\mathrm{var}(g_j) \subseteq \underline{X}(I_l), l \in [p]$.

Next, with $l \in [p]$ fixed, d a positive integer and $g \in \mathbb{R}\langle \underline{X}(I_l)\rangle$, let $M_d(\mathbf{y}, I_l)$ (resp. $M_d(g\mathbf{y}, I_l)$) be the moment (resp. localizing) submatrix obtained from $M_d(\mathbf{y})$ (resp. $M_d(g\mathbf{y})$) by retaining only those rows (and columns) indexed by $w \in \langle \underline{X}(I_l)\rangle$ of $M_d(\mathbf{y})$ (resp. $M_d(g\mathbf{y})$).

Then with $\hat{d} \geq d := \max\{\lceil \deg(f)/2\rceil, \lceil \deg(g_1)/2\rceil, \ldots, \lceil \deg(g_m)/2\rceil\}$, the moment SDP relaxation for (EQ_0) based on correlative sparsity is defined as

$$\left(\mathrm{EQ}_{\hat{d}}^{\mathrm{cs}}\right): \begin{array}{rl} \lambda_{\hat{d}}^{\mathrm{cs}}(f,S) := & \inf\ L_{\mathbf{y}}(f) \\ & \text{s.t. } M_{\hat{d}}(\mathbf{y}, I_l) \succeq 0, \quad l \in [p], \\ & M_{\hat{d}-d_j}(g_j\mathbf{y}, I_l) \succeq 0, \quad j \in J_l, l \in [p], \\ & y_1 = 1. \end{array} \tag{35}$$

Remark 3 As shown in [21], under Archimedean's condition (slightly stronger than compactness), the sequence $(\lambda_{\hat{d}}^{\mathrm{cs}}(f,S))_{\hat{d}\geq d}$ converges to the global optimum $\lambda_{\min}(f,S)$.

4.2 Eigenvalue optimization for noncommutative polynomials with combined correlative-term sparsity

The combination of correlative sparsity and term sparsity proceeds in a similar manner as for the commutative case in [48]. Assume that $f = \sum_w a_w w \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$ and $\underline{S} = \{g_1, \ldots, g_m\} \subseteq \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle$, G^{csp} is the csp graph associated with f and $\underline{S}$, and $\overline{G}^{\mathrm{csp}}$ is a chordal extension of G^{csp}. Let $I_l, l \in [p]$ be the maximal cliques of $\overline{G}^{\mathrm{csp}}$ with cardinality denoted by $n_l, l \in [p]$. Then the set of variables $\underline{X}$ is partitioned into $\underline{X}(I_1), \underline{X}(I_2), \ldots, \underline{X}(I_p)$. Let $J_1, \ldots, J_p$ be defined as in Sect. 4.1.

Now we consider the term sparsity pattern for each subsystem involving the variables $\underline{X}(I_l), l \in [p]$ respectively as follows. Let

$$\mathbf{A} := \mathrm{supp}(f) \cup \bigcup_{j=1}^{m} \mathrm{supp}(g_j) \text{ and } \mathbf{A}_l := \left\{w \in \mathbf{A} \mid \mathrm{var}(w) \subseteq \underline{X}(I_l)\right\}, \tag{36}$$

for $l \in [p]$. As before, let $g_0 = 1$, $d_j = \lceil \deg(g_j)/2\rceil$, $j \in \{0\} \cup [m]$ and $d = \max\{\lceil \deg(f)/2\rceil, d_1, \ldots, d_m\}$. Fix a relaxation order $\hat{d} \geq d$. Let $\mathbf{W}_{\hat{d}-d_j,l}$ be the standard monomial basis of degree $\leq \hat{d} - d_j$ with respect to the variables $\underline{X}(I_l)$ and $G_{\hat{d},l}^{\mathrm{tsp}}$ be the tsp graph with nodes $\mathbf{W}_{\hat{d},l}$ associated with $\mathbf{A}_l$ defined as in Sect. 3.2. Assume that $G_{\hat{d},l,0}^{(0)} = G_{\hat{d},l}^{\mathrm{tsp}}$ and $G_{\hat{d},l,j}^{(0)}, j \in J_l, l \in [p]$ are empty graphs. Letting

$$\mathscr{C}_{\hat{d}}^{\ (k-1)} := \bigcup_{l=1}^{p} \bigcup_{j\in\{0\}\cup J_l} \operatorname{supp}_{g_j}\Big(G_{\hat{d},l,j}^{(k-1)}\Big) \cup \mathbf{W}_{\hat{d}}^2, \quad k \geq 1, \tag{37}$$

we iteratively define a sequence of graphs $(G_{\hat{d},l,j}^{(k)}(V_{\hat{d},l,j}, E_{\hat{d},l,j}^{(k)}))_{k\geq 1}$ with $V_{\hat{d},l,j} = \mathbf{W}_{\hat{d}-d_j,l}$ for each $j \in \{0\}\cup J_l, l \in [p]$ by

$$G_{\hat{d},l,j}^{(k)} := \overline{F_{\hat{d},l,j}^{(k)}}, \tag{38}$$

where $F_{\hat{d},l,j}^{(k)}$ is the graph with $V(F_{\hat{d},l,j}^{(k)}) = \mathbf{W}_{\hat{d}-d_j,l}$ and

$$E\Big(F_{\hat{d},l,j}^{(k)}\Big) = \Big\{\{u,v\} \mid u \neq v \in \mathbf{W}_{\hat{d}-d_j,l}, u^\star \operatorname{supp}(g_j) v \cap \mathscr{C}_{\hat{d}}^{(k-1)} \neq \emptyset\Big\}. \tag{39}$$

Let $r_{l,j} = |\mathbf{W}_{\hat{d}-d_j,l}|$ for all l, j. Then for each $k \geq 1$ (called the *sparse order*), the sparse moment relaxation based on combined correlative-term sparsity for (EQ_0) is given by

$$\Big(\mathrm{EQ}_{\hat{d},k}^{\text{cs-ts}}\Big): \begin{aligned} \lambda_{\hat{d},k}^{\text{cs-ts}}(f,S) := \ & \inf\ L_{\mathbf{y}}(f) \\ & \text{s.t. } B_{G_{\hat{d},l,0}^{(k)}} \circ M_{\hat{d}}(\mathbf{y}, I_l) \in \Pi_{G_{\hat{d},l,0}^{(k)}}\Big(\mathbb{S}_{r_{l,0}}^+\Big), l \in [p], \\ & \quad B_{G_{\hat{d},l,j}^{(k)}} \circ M_{\hat{d}-d_j}\big(g_j\mathbf{y}, I_l\big) \in \Pi_{G_{\hat{d},l,j}^{(k)}}\Big(\mathbb{S}_{r_{l,j}}^+\Big), j \in J_l, l \in [p], \\ & \quad y_1 = 1. \end{aligned} \tag{40}$$

For any l, j, write $M_{\hat{d}-d_j}(g_j\mathbf{y}, I_l) = \sum_w D_w^{l,j} y_w$ for appropriate matrices $\{D_w^{l,j}\}_{l,j,w}$. Then for each $k \geq 1$, the dual of $(\mathrm{EQ}_{\hat{d},k}^{\text{cs-ts}})$ reads as

$$\begin{aligned} \sup\ & \lambda \\ \text{s.t.}\ & \sum_{l=1}^{p} \sum_{j\in\{0\}\cup J_l} \langle Q_{l,j}, D_w^{l,j}\rangle + \lambda\delta_{1w} = a_w, \quad \forall w \in \mathscr{C}_{\hat{d}}^{(k)}, \\ & Q_{l,j} \in \mathbb{S}_{r_{l,j}}^+ \cap \mathbb{S}_{G_{\hat{d},l,j}^{(k)}}, \quad j \in \{0\}\cup J_l, l \in [p], \end{aligned} \tag{41}$$

where $\mathscr{C}_{\hat{d}}^{(k)}$ is defined as in (37).

By similar arguments as for Theorem 6, we can prove the following theorem.

Theorem 7 *Assume that* $\{f\} \cup S \subseteq \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle$. *Then the following hold*:

(1) *Fixing a relaxation order* $\hat{d} \geq d$, *the sequence* $(\lambda_{\hat{d},k}^{\text{cs-ts}}(f,S))_{k\geq 1}$ *is monotone nondecreasing and* $\lambda_{\hat{d},k}^{\text{cs-ts}}(f,S) \leq \lambda_{\hat{d}}^{\text{cs}}(f,S)$ *for all* $k \geq 1$ (*with* $\lambda_{\hat{d}}^{\text{cs}}(f,S)$ *defined in Sect.* 4.1).
(2) *Fixing a sparse order* $k \geq 1$, *the sequence* $(\lambda_{\hat{d},k}^{\text{cs-ts}}(f,S))_{\hat{d}\geq d}$ *is monotone nondecreasing*.
(3) *If the maximal chordal extension is used in* (38), *then* $(\lambda_{\hat{d},k}^{\text{cs-ts}}(f,S))_{k\geq 1}$ *converges to* $\lambda_{\hat{d}}^{\text{cs}}(f,S)$ *in finitely many steps*.

From Theorem 7, we deduce the following two-level hierarchy of lower bounds for the optimum $\lambda_{\min}(f,S)$ of (EQ_0):

$$\begin{array}{ccccccc}
\lambda_{d,1}^{\text{cs-ts}}(f,S) & \le & \lambda_{d,2}^{\text{cs-ts}}(f,S) & \le \cdots \le & \lambda_{d}^{\text{cs}}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\lambda_{d+1,1}^{\text{cs-ts}}(f,S) & \le & \lambda_{d+1,2}^{\text{cs-ts}}(f,S) & \le \cdots \le & \lambda_{d+1}^{\text{cs}}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\vdots & & \vdots & \vdots & \vdots \\
\wedge| & & \wedge| & & \wedge| \\
\lambda_{\hat{d},1}^{\text{cs-ts}}(f,S) & \le & \lambda_{\hat{d},2}^{\text{cs-ts}}(f,S) & \le \cdots \le & \lambda_{\hat{d}}^{\text{cs}}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\vdots & & \vdots & \vdots & \vdots
\end{array} \tag{42}$$

5 Trace optimization for noncommutative polynomials with term sparsity

The results presented in the previous sections concerning eigenvalue optimization for noncommutative polynomials with term sparsity can be slightly adjusted to deal with trace optimization for noncommutative polynomials with term sparsity. We present the main results concerning trace optimization in this section and omit the proofs.

5.1 The unconstrained case

Let $f = \sum_{w\in\mathbf{A}} a_w w \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle$ with $\operatorname{supp}(f) = \mathbf{A}$ (w.l.o.g. assuming $1 \in \mathbf{A}$) and let $d = \operatorname{cdeg}(f)$. We define $H_0(V, E_0)$ to be the graph with $V = \mathbf{W}_d$ and

$$E_0 = \big\{\{u,v\} \mid (u,v) \in V\times V,\ u \neq v,\ \big[u^\star v\big] \in \big[\mathbf{A} \cup \mathbf{W}_d^2\big]\big\}. \tag{43}$$

We iteratively define a sequence of graphs $(H_k(V,E_k))_{k\ge 1}$ by

$$H_k := \overline{\mathrm{CSE}\big(H_{k-1}\big)}, \tag{44}$$

where $\mathrm{CSE}(H_{k-1})$ (the cyclic support extension of H_{k-1}) is the graph with nodes $\mathbf{W}_d$ and with edges

$$E\big(\mathrm{CSE}(H_{k-1})\big) := \big\{\{u,v\} \mid (u,v) \in V\times V,\ u \neq v,\ \big[u^\star v\big] \in \big[\operatorname{supp}(H_{k-1}) \cup \mathbf{W}_d^2\big]\big\}.$$

Let $r = |\mathbf{W}_d|$. As for eigenvalue optimization, we can consider the following series of sparse moment relaxations for (TP) indexed by $k \ge 1$:

$$\left(\mathrm{TP}^k\right): \quad \begin{aligned} \mu_k(f) := \inf\ & L_{\mathbf{y}}(f) \\ \text{s.t.}\ & B_{H_k} \circ M_d(\mathbf{y}) \in \Pi_{H_k}(\mathbb{S}_r^+), \\ & \left[B_{H_k} \circ M_d(\mathbf{y})\right]_{uv} = \left[B_{H_k} \circ M_d(\mathbf{y})\right]_{wz}, \text{for all } u^\star v \overset{\text{cyc}}{\sim} w^\star z, \\ & y_1 = 1. \end{aligned} \tag{45}$$

The dual of (TP^k) reads as

$$\begin{aligned} \sup\ & \mu \\ \text{s.t.}\ & \sum_{w \overset{\text{cyc}}{\sim} v} \left(\langle Q, A_w \rangle + \mu \delta_{1w}\right) = \sum_{w \overset{\text{cyc}}{\sim} v} a_w, \quad \forall v \in \left[\mathrm{supp}(H_k) \cup \mathbf{W}_d^2\right], \\ & Q \in \mathbb{S}_r^+ \cap \mathbb{S}_{H_k}, \end{aligned} \tag{46}$$

where A_w is defined as in Sect. 2.5. We call k the *sparse order*. There is no duality gap between (TP^k) and its dual. By construction, one has $H_k \subseteq H_{k+1}$ for all $k \geq 1$ and therefore the sequence of graphs $(H_k(V, E_k))_{k \geq 1}$ stabilizes after a finite number of steps. We denote the stabilized graph by $H_\bullet(V, E_\bullet)$ and the optimum of the corresponding SDP relaxation by $\mu_\bullet(f)$.

As for eigenvalue optimization, we obtain the following hierarchy of lower bounds for $\mathrm{tr}_{\min}(f)$:

$$\mu_1(f) \leq \mu_2(f) \leq \cdots \leq \mu_\bullet(f) \leq \mu(f) \leq \mathrm{tr}_{\min}(f). \tag{47}$$

Moreover, if the maximal chordal extension is used in (44), then $(\mu_k(f))_{k \geq 1}$ converges to $\mu(f)$ in finitely many steps, i.e., $\mu_\bullet(f) = \mu(f)$.

Remark 4 The monomial basis $\mathbf{W}_d$ used in this subsection can be replaced by the reduced monomial basis returned by the tracial Newton polytope method [11, §3.3]. However, the implementation of this method involves solving $\binom{n+d}{d}$ LPs with $n+1$ decision variables, which becomes expensive for large n. Hence we stick to the standard monomial basis $\mathbf{W}_d$ for the numerical experiments performed in this paper (see Sect. 6).

5.2 The constrained case

Assume that $f = \sum_w a_w w \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X} \rangle$ and $S = \{g_1, \ldots, g_m\} \subseteq \mathrm{Sym}\,\mathbb{R}\langle \underline{X} \rangle$. As before, let $\mathbf{A} = \mathrm{supp}(f) \cup \bigcup_{j=1}^m \mathrm{supp}(g_j)$, $g_0 = 1$ and $d_j = \lceil \deg(g_j)/2 \rceil$, $j \in \{0\} \cup [m]$. Let $d = \max\{\lceil \mathrm{cdeg}(f)/2 \rceil, d_1, \ldots, d_m\}$. Fix a relaxation order $\hat{d} \geq d$. We define a graph $H_{\hat{d}}^{\mathrm{tsp}}(V_{\hat{d}}, E_{\hat{d}}^{\mathrm{tsp}})$ with $V_{\hat{d}} = \mathbf{W}_{\hat{d}}$ and

$$E_{\hat{d}}^{\mathrm{tsp}} = \left\{ \{u, v\} \mid (u, v) \in V_{\hat{d}} \times V_{\hat{d}},\ u \neq v,\ \left[u^\star v\right] \in \left[\mathbf{A} \cup \mathbf{W}_{\hat{d}}^2\right] \right\}, \tag{48}$$

which is called the *cyclic tsp graph* associated with $\mathbf{A}$ (or f and S). Let $H_{\hat{d},0}^{(0)} = H_{\hat{d}}^{\mathrm{tsp}}$ and $H_{\hat{d},j}^{(0)}$ be the empty graph for $j \in [m]$. We iteratively define a sequence of graphs

$(H_{\hat{d},j}^{(k)}(V_{\hat{d},j}, E_{\hat{d},j}^{(k)}))_{k\geq 1}$ with $V_{\hat{d},j} = \mathbf{W}_{\hat{d}-d_j}$ for each $j \in \{0\} \cup [m]$ via two successive operations:

(1) *cyclic support extension* Let $K_{\hat{d},j}^{(k)}$ be the graph with $V(K_{j,\hat{d}}^{(k)}) = \mathbf{W}_{\hat{d}-d_j}$ and

$$\begin{aligned} E\left(K_{\hat{d},j}^{(k)}\right) =& \{\{u,v\} \mid (u,v) \in \mathbf{W}_{\hat{d}-d_j} \times \mathbf{W}_{\hat{d}-d_j},\ u \neq v, \\ & \left[u^\star \operatorname{supp}(g_j) v\right] \cap \left[\bigcup_{j=0}^{m} \operatorname{supp}_{g_j}\left(H_{j,\hat{d}}^{(k-1)}\right) \cup \mathbf{W}_{\hat{d}}^2\right] \neq \emptyset\}. \end{aligned} \tag{49}$$

(2) *chordal extension* Let

$$H_{\hat{d},j}^{(k)} := \overline{K_{\hat{d},j}^{(k)}}. \tag{50}$$

Let $r_j = |\mathbf{W}_{\hat{d}-d_j}|$. As for eigenvalue optimization, we then consider the following series of sparse moment relaxations for $(\mathrm{TQ}_{\hat{d}})$ indexed by $k \geq 1$:

$$\left(\mathrm{TQ}_{\hat{d},k}^{\mathrm{ts}}\right): \begin{aligned} \mu_{\hat{d},k}^{\mathrm{ts}}(f,S) :=\ & \inf\ L_{\mathbf{y}}(f) \\ & \text{s.t. } B_{H_{\hat{d},0}^{(k)}} \circ M_{\hat{d}}(\mathbf{y}) \in \Pi_{H_{\hat{d},0}^{(k)}}\left(\mathbb{S}_{r_0}^{+}\right), \\ & \quad B_{H_{\hat{d},j}^{(k)}} \circ M_{\hat{d}-d_j}\left(g_j \mathbf{y}\right) \in \Pi_{H_{\hat{d},j}^{(k)}}\left(\mathbb{S}_{r_j}^{+}\right), j \in [m], \\ & \quad \left[B_{H_{\hat{d},0}^{(k)}} \circ M_{\hat{d}}(\mathbf{y})\right]_{uv} = \left[B_{H_{\hat{d},0}^{(k)}} \circ M_{\hat{d}}(\mathbf{y})\right]_{wz}, \text{for all } u^\star v \overset{\mathrm{cyc}}{\sim} w^\star z, \\ & \quad y_1 = 1. \end{aligned} \tag{51}$$

We call k the *sparse order*. By construction, one has $H_{\hat{d},j}^{(k)} \subseteq H_{\hat{d},j}^{(k+1)}$ for all j, k. Therefore, for every j, the sequence of graphs $(H_{\hat{d},j}^{(k)})_{k\geq 1}$ stabilizes after a finite number of steps. We denote the stabilized graphs by $H_{\hat{d},j}^{(\bullet)}$ for all j and the optimum of the corresponding SDP relaxation by $\mu_{\hat{d},\bullet}(f,S)$.

For each $k \geq 1$, the dual of $(\mathrm{TQ}_{\hat{d},k}^{\mathrm{ts}})$ reads as

$$\begin{aligned} \sup\ & \mu \\ \text{s.t. }\ & \textstyle\sum_{w \overset{\mathrm{cyc}}{\sim} v} \left(\sum_{j=0}^{m} \langle Q_j, D_w^j \rangle + \mu \delta_{1w}\right) = \sum_{w \overset{\mathrm{cyc}}{\sim} v} a_w, \\ & \qquad\qquad \forall v \in [\textstyle\bigcup_{j=0}^{m} \operatorname{supp}_{g_j}(H_{\hat{d},j}^{(k)})) \cup \mathbf{W}_{\hat{d}}^2], \\ & Q_j \in \mathbb{S}_{r_j}^{+} \cap \mathbb{S}_{H_{\hat{d},j}^{(k)}}, \quad j \in \{0\} \cup [m], \end{aligned} \tag{52}$$

where D_w^j is defined as in Sect. 2.5.

As for eigenvalue optimization, we have the following conclusions.

Theorem 8 *Assume that* $\{f\} \cup S \in \operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle$. *Then the following hold*:

(1) *If* $\mathcal{D}_S$ *contains an nc* ε*-neighborhood of* 0, *then for all* $\hat{d}, k$, *there is no duality gap between* $(\mathrm{TQ}_{\hat{d},k}^{\mathrm{ts}})$ (51) *and its dual* (52).

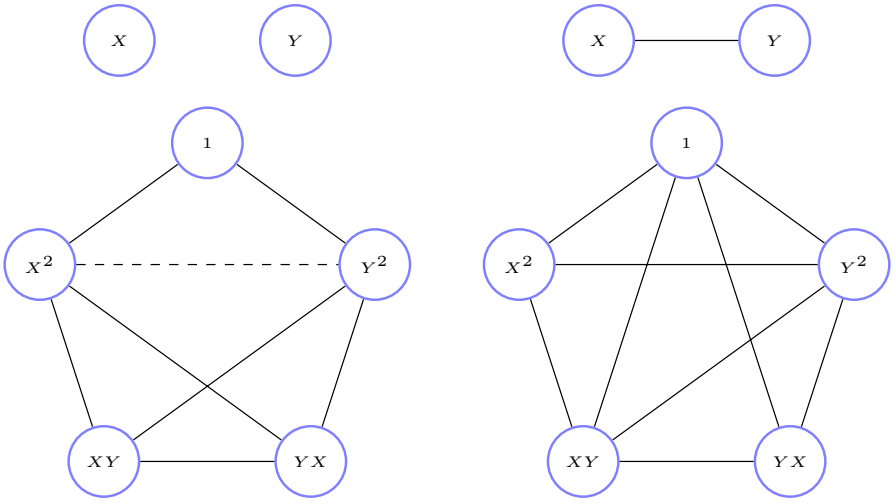

Fig. 5 The graph sequence $(H_{2,0}^{(k)})_{k\geq 1}$ in Example 5: left for $k = 1$; right for $k = 2$. The dashed edges are added after a chordal extension

(2) *Fixing a relaxation order* $\hat{d} \geq d$, *the sequence* $(\mu_{\hat{d},k}^{\mathrm{ts}}(f,S))_{k\geq 1}$ *is monotone nondecreasing and* $\mu_{\hat{d},k}^{\mathrm{ts}}(f,S) \leq \mu_{\hat{d}}(f,S)$ *for all* $k \geq 1$ (*with* $\mu_{\hat{d}}(f,S)$ *defined in* (15)).

(3) *Fixing a sparse order* $k \geq 1$, *the sequence* $(\mu_{\hat{d},k}^{\mathrm{ts}}(f,S))_{\hat{d}\geq d}$ *is monotone nondecreasing*.

(4) *If the maximal chordal extension is used in* (50), *then* $(\mu_{\hat{d},k}^{\mathrm{ts}}(f,S))_{k\geq 1}$ *converges to* $\mu_{\hat{d}}(f,S)$ *in finitely many steps, i.e.,* $\mu_{\hat{d},\bullet}(f,S) = \mu_{\hat{d}}(f,S)$.

Following from Theorem 8, we have the following two-level hierarchy of lower bounds for the optimum $\mathrm{tr}_{\min}(f,S)^{\mathrm{II}_1}$:

$$\begin{array}{ccccccc}
\mu_{d,1}^{\mathrm{ts}}(f,S) & \leq & \mu_{d,2}^{\mathrm{ts}}(f,S) & \leq \cdots \leq & \mu_{d}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\mu_{d+1,1}^{\mathrm{ts}}(f,S) & \leq & \mu_{d+1,2}^{\mathrm{ts}}(f,S) & \leq \cdots \leq & \mu_{d+1}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\vdots & & \vdots & \vdots & \vdots \\
\wedge| & & \wedge| & & \wedge| \\
\mu_{\hat{d},1}^{\mathrm{ts}}(f,S) & \leq & \mu_{\hat{d},2}^{\mathrm{ts}}(f,S) & \leq \cdots \leq & \mu_{\hat{d}}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\vdots & & \vdots & \vdots & \vdots
\end{array} \tag{53}$$

The array of lower bounds (53) (and its associated moment-SOHS relaxations (51)-(52)) is what we call the NCTSSOS hierarchy associated with (TQ_0).

Example 5 Consider again $f = 2 - X^2 + XY^2X - Y^2 + XYXY + YXYX + X^3Y + YX^3 + XY^3 + Y^3X$ and $S = \{1 - X^2, 1 - Y^2\}$. The graph sequence $(H_{2,0}^{(k)})_{k\geq 1}$ for f and S is given in Fig. 5. In fact the graph sequence $(H_{2,j}^{(k)})_{k\geq 1}$ stabilizes at $k = 2$ for all j (with approximately smallest chordal extensions). Using `NCTSSOS`, we obtain that $\mu_{2,1}^{\mathrm{ts}}(f, S) = \mu_{2,2}^{\mathrm{ts}}(f, S) = \mu_2(f, S) \approx -2$.

5.3 Combining correlative sparsity with term sparsity

We can also combine correlative sparsity with term sparsity for trace optimization. Fix a relaxation order $\hat{d} \geq d$. Let $\mathbf{A}, I_l, J_l, \mathbf{A}_l, \mathbf{W}_{\hat{d},l}, \mathbf{W}_{\hat{d}-d_j,l}$ be defined as in Sect. 4.2. Let $H_{\hat{d},l}^{\mathrm{tsp}}$ be the cyclic tsp graph with nodes $\mathbf{W}_{\hat{d},l}$ associated with $\mathbf{A}_l$ defined as in Sect. 5.2. Assume that $H_{\hat{d},l,0}^{(0)} = H_{\hat{d},l}^{\mathrm{tsp}}$ and $H_{\hat{d},l,j}^{(0)}, j \in J_l, l \in [p]$ are empty graphs. Letting

$$\mathscr{D}_{\hat{d}}^{(k-1)} := \bigcup_{l=1}^{p} \bigcup_{j\in\{0\}\cup J_l} \mathrm{supp}_{g_j}\left(H_{\hat{d},l,j}^{(k-1)}\right) \cup \mathbf{W}_{\hat{d}}^2, \quad k \geq 1, \tag{54}$$

we iteratively define a sequence of graphs $(H_{\hat{d},l,j}^{(k)}(V_{\hat{d},l,j}, E_{\hat{d},l,j}^{(k)}))_{k\geq 1}$ with $V_{\hat{d},l,j} = \mathbf{W}_{\hat{d}-d_j,l}$ for each $j \in \{0\} \cup J_l, l \in [p]$ by

$$H_{\hat{d},l,j}^{(k)} := \overline{K_{\hat{d},l,j}^{(k)}}, \tag{55}$$

where $K_{\hat{d},l,j}^{(k)}$ is the graph with $V(K_{\hat{d},l,j}^{(k)}) = \mathbf{W}_{\hat{d}-d_j,l}$ and

$$E\left(K_{\hat{d},l,j}^{(k)}\right) = \left\{ \{u, v\} \mid u \neq v \in \mathbf{W}_{\hat{d}-d_j,l}, \left[u^\star \mathrm{supp}(g_j) v\right] \cap \left[\mathscr{D}_{\hat{d}}^{(k-1)}\right] \neq \emptyset \right\}. \tag{56}$$

Let $r_{l,j} = |\mathbf{W}_{\hat{d}-d_j,l}|$ for all l, j. Then for each $k \geq 1$, the moment relaxation based on combined correlative-term sparsity for (TQ_0) is given by

$$\left(\mathrm{TQ}_{\hat{d},k}^{\text{cs-ts}}\right): \begin{aligned} \mu_{\hat{d},k}^{\text{cs-ts}}(f, S) := \ & \inf\ L_{\mathbf{y}}(f) \\ & \text{s.t. } B_{H_{\hat{d},l,0}^{(k)}} \circ M_{\hat{d}}(\mathbf{y}, I_l) \in \Pi_{H_{\hat{d},l,0}^{(k)}}(\mathbb{S}_{r_{l,0}}^+), l \in [p], \\ & \quad B_{H_{\hat{d},l,j}^{(k)}} \circ M_{\hat{d}-d_j}(g_j\mathbf{y}, I_l) \in \Pi_{H_{\hat{d},l,j}^{(k)}}(\mathbb{S}_{r_{l,j}}^+), j \in J_l, l \in [p], \\ & \quad [B_{H_{\hat{d},l,0}^{(k)}} \circ M_{\hat{d}}(\mathbf{y}, I_l)]_{uv} = [B_{H_{\hat{d},l,0}^{(k)}} \circ M_{\hat{d}}(\mathbf{y}, I_l)]_{wz}, \\ & \qquad \text{for all } u^\star v \overset{\mathrm{cyc}}{\sim} w^\star z, l \in [p], \\ & \quad y_1 = 1. \end{aligned} \tag{57}$$

For each $k \geq 1$, the dual of $(\mathrm{TQ}_{\hat{d},k}^{\text{cs-ts}})$ reads as

$$\begin{aligned} \sup\ & \mu \\ \text{s.t. } & \textstyle\sum_{w \overset{\mathrm{cyc}}{\sim} v} \left(\sum_{l=1}^{p} \sum_{j\in\{0\}\cup J_l} \langle Q_{l,j}, D_w^{l,j}\rangle + \mu\delta_{1w}\right) = \sum_{w \overset{\mathrm{cyc}}{\sim} v} a_w, \forall v \in [\mathscr{D}_{\hat{d}}^{(k)}], \\ & Q_{l,j} \in \mathbb{S}_{r_{l,j}}^+ \cap \mathbb{S}_{H_{\hat{d},l,j}^{(k)}}, j \in \{0\} \cup J_l, l \in [p], \end{aligned} \tag{58}$$

where $D_w^{l,j}$ is defined as in Sect. 4.2 and $\mathscr{D}_{\hat{d}}^{(k)}$ is defined as in (54).

Theorem 9 *Assume that* $\{f\} \cup S \subseteq \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle$. *Let* $\mu_{\hat{d}}^{\text{cs}}(f,S)$ *be the optimum of the* $\hat{d}$*-th order sparse moment relaxation based on correlative sparsity for* (TQ_0). *Then the following hold*:

(1) *Fixing a relaxation order* $\hat{d} \geq d$, *the sequence* $(\mu_{\hat{d},k}^{\text{cs-ts}}(f,S))_{k\geq 1}$ *is monotone nondecreasing and* $\mu_{\hat{d},k}^{\text{cs-ts}}(f,S) \leq \mu_{\hat{d}}^{\text{cs}}(f,S)$ *for all* $k \geq 1$.
(2) *Fixing a sparse order* $k \geq 1$, *the sequence* $(\mu_{\hat{d},k}^{\text{cs-ts}}(f,S))_{\hat{d}\geq d}$ *is monotone nondecreasing*.
(3) *If the maximal chordal extension is used in* (55), *then* $(\mu_{\hat{d},k}^{\text{cs-ts}}(f,S))_{k\geq 1}$ *converges to* $\mu_{\hat{d}}^{\text{cs}}(f,S)$ *in finitely many steps*.

From Theorem 9, we deduce the following two-level hierarchy of lower bounds for the optimum $\text{tr}_{\min}(f,S)^{\text{II}_1}$:

$$\begin{array}{ccccccc}
\mu_{d,1}^{\text{cs-ts}}(f,S) & \leq & \mu_{d,2}^{\text{cs-ts}}(f,S) & \leq \cdots \leq & \mu_{d}^{\text{cs}}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\mu_{d+1,1}^{\text{cs-ts}}(f,S) & \leq & \mu_{d+1,2}^{\text{cs-ts}}(f,S) & \leq \cdots \leq & \mu_{d+1}^{\text{cs}}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\vdots & & \vdots & \vdots & \vdots \\
\wedge| & & \wedge| & & \wedge| \\
\mu_{\hat{d},1}^{\text{cs-ts}}(f,S) & \leq & \mu_{\hat{d},2}^{\text{cs-ts}}(f,S) & \leq \cdots \leq & \mu_{\hat{d}}^{\text{cs}}(f,S) \\
\wedge| & & \wedge| & & \wedge| \\
\vdots & & \vdots & \vdots & \vdots
\end{array} \tag{59}$$

6 Numerical experiments

In this section, we present numerical results of the proposed NCTSSOS hierarchies for both unconstrained and constrained noncommutative polynomial optimization problems. Our tool to implement these hierarchies, named `NCTSSOS`, is written as a Julia package. `NCTSSOS` utilizes the Julia packages LightGraphs [7] to handle graphs, ChordalGraph [43] to generate approximately smallest chordal extensions and JuMP [14] to model SDP. Finally, `NCTSSOS` relies on MOSEK [3] to solve SDP. `NCTSSOS` is freely available at

https://github.com/wangjie212/NCTSSOS.

All numerical examples were computed on an Intel Core i5-8265U@1.60GHz CPU with 8GB RAM memory. The timing includes the time for pre-processing (to get the block structure in `NCTSSOS`), the time for modeling SDP and the time for solving SDP. For most examples tested in this paper (except the random ones), we use approximately smallest chordal extensions since they are much more scalable than the maximal chordal extension while yielding (almost) the

Table 1 The notation

n	The number of nc variables
$\hat{d}$	The relaxation order
k	The sparse order
t	The number of terms
mb	The maximal size of blocks
opt	The optimal value
Time	Running time in seconds
0	A number with absolute value less than $1\times^{-4}$
–	MOSEK running out of memory

same optimum. For comparison purpose, we also implement the dense moment-SOHS relaxation in `NCTSSOS`. The notations that we use in the following subsections are listed in Table 1.

6.1 Eigenvalue optimization examples

We first focus on the unconstrained case and consider the eigenvalue minimization problem for the following functions.

- The nc version of the Broyden banded function

$$f_{\mathrm{Bb}}(\mathbf{x}) = \sum_{i=1}^{n} \left(2X_i + 5X_i^3 + 1 - \sum_{j\in J_i}\left(X_j + X_j^2\right)\right)^{\star} \left(2X_i + 5X_i^3 + 1 - \sum_{j\in J_i}\left(X_j + X_j^2\right)\right),$$

  where $J_i = \{j \mid j \neq i, \max(1, i-5) \leq j \leq \min(n, i+1)\}$.
- The nc version of the chained singular function

$$\begin{aligned} f_{\mathrm{cs}}(\mathbf{x}) = \sum_{i\in J} &\big(\big(X_i + 10X_{i+1}\big)^{\star}\big(X_i + 10X_{i+1}\big) + 5\big(X_{i+2} - X_{i+3}\big)^{\star}\big(X_{i+2} - X_{i+3}\big) \\ &+ \big(X_{i+1}^2 - 4X_{i+1}X_{i+2} + 4X_{i+2}^2\big)^{\star}\big(X_{i+1}^2 - 4X_{i+1}X_{i+2} + 4X_{i+2}^2\big) \\ &+ 10\big(X_i^2 - 20X_iX_{i+3} + 100X_{i+3}^2\big)^{\star}\big(X_i^2 - 20X_iX_{i+3} + 100X_{i+3}^2\big), \end{aligned}$$

  where $J = \{1, 3, 5, \ldots, n-3\}$.
- The nc version of the generalized Rosenbrock function

$$f_{\mathrm{gR}}(\mathbf{x}) = 1 + \sum_{i=1}^{n} \Big(100\big(X_i - X_{i-1}^2\big)^{\star}\big(X_i - X_{i-1}^2\big) + \big(1 - X_i\big)^{\star}\big(1 - X_i\big)\Big).$$

- The nc version of the chained Wood function

Table 2 The eigenvalue minimization for the nc Broyden banded function with $\hat{d} = 3, k = 1$

n	Sparse			Dense		
	mb	opt	time	mb	opt	Time
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	–	–	–

$$\begin{aligned} f_{\mathrm{cW}}(\mathbf{x}) =& 1 + \sum_{i\in J} \Big(100\big(X_{i+1} - X_i^2\big)^\star \big(X_{i+1} - X_i^2\big) + \big(1 - X_i\big)^\star \big(1 - X_i\big) + 90\big(X_{i+3} \\ & -X_{i+2}^2\big)^\star \big(X_{i+3} - X_{i+2}^2\big) + \big(1 - X_{i+2}\big)^\star \big(1 - X_{i+2}\big) + 10\big(X_{i+1} + X_{i+3} \\ & -2\big)^\star \big(X_{i+1} + X_{i+3} - 2\big) + 0.1\big(X_{i+1} - X_{i+3}\big)^\star \big(X_{i+1} - X_{i+3}\big) \Big), \end{aligned}$$

where $J = \{1, 3, 5, \ldots, n-3\}$ and $4|n$.

- The nc version of the Broyden tridiagonal function

$$\begin{aligned} f_{\mathrm{Bt}}(\mathbf{x}) =& \big(3X_1 - 2X_1^2 - 2X_2 + 1\big)^\star \big(3X_1 - 2X_1^2 - 2X_2 + 1\big) \\ & + \sum_{i=2}^{n-1} \big(3X_i - 2X_i^2 - X_{i-1} - 2X_{i+1} + 1\big)^\star \big(3X_i - 2X_i^2 - X_{i-1} - 2X_{i+1} + 1\big) \\ & + \big(3X_n - 2X_n^2 - X_{n-1} + 1\big)^\star \big(3X_n - 2X_n^2 - X_{n-1} + 1\big). \end{aligned}$$

To solve the unconstrained eigenvalue minimization problem of these functions, we always rely on the Newton chip method to compute a monomial basis, which turns out to be much smaller than the standard monomial basis. We compute $\lambda_1^{\mathrm{ts}}(f)$ (the optimum of $(\mathrm{EQ}_1^{\mathrm{ts}})$) using approximately smallest chordal extensions and compare the resulting values with $\lambda_{\min}(f)$ (the optimum of (EP)) corresponding to the dense approach. The results are reported in Table 2, 3, 4, 5, 6, respectively. It is evident from these tables that our sparse approach is much more scalable than the dense approach. The dense approach can never be executed due to the memory limit when the problem has over 100 variables while the sparse approach can easily handle problems with 4000 variables. Meanwhile when the dense approach is executable,

Table 3 The eigenvalue minimization for the nc chained singular function with $\hat{d} = 2, k = 1$

n	sparse			dense		
	mb	opt	time	mb	opt	Time
20	3	−0.0004	0.06	59	−0.0001	1.65
40	3	−0.0024	0.10	119	−0.0003	54.0
60	3	0	0.16	179	−0.0002	516
80	3	−0.0005	0.19	–	–	–
100	3	0	0.20	–	–	–
200	3	−0.0001	0.50	–	–	–
400	3	−0.0331	0.97	–	-	–
600	3	−0.0005	1.85	–	–	–
800	3	−0.0381	2.69	–	–	–
1000	3	−0.0074	4.10	–	–	–
2000	3	−0.0004	15.7	–	–	–
3000	3	−0.0065	32.4	–	–	–
4000	3	−0.0007	58.7	–	–	–

Table 4 The eigenvalue minimization for the nc generalized Rosenbrock function with $\hat{d} = 2, k = 1$

n	Sparse			Dense		
	mb	opt	Time	mb	opt	Time
20	3	1.0000	0.06	40	1.0000	0.33
40	3	1.0000	0.06	80	1.0000	4.59
60	3	1.0000	0.07	120	1.0000	31.9
80	3	1.0000	0.08	160	1.0000	151
100	3	1.0000	0.08	200	1.0000	557
200	3	0.9999	0.15	–	–	–
400	3	0.9999	0.45	–	–	–
600	3	0.9999	0.70	–	–	–
800	3	0.9999	1.03	–	–	–
1000	3	0.9998	1.38	–	–	–
2000	3	1.0000	4.76	–	–	–
3000	3	1.0000	10.7	–	–	–
4000	3	0.9999	18.9	–	–	–

the optimal value provided by the sparse approach is quite close (or even equal in many cases) to the one provided by the dense approach.

Now let us consider the constrained case. Let $\mathcal{D}$ be the semialgebraic set defined by $\{1 - X_1^2, \ldots, 1 - X_n^2, X_1 - 1/3, \ldots, X_n - 1/3\}$ and the optimization problem is minimizing the eigenvalue of the nc Broyden banded function over $\mathcal{D}$. We compute $\lambda_{\hat{d},1}^{\text{cs-ts}}(f, S)$ (the optimum of $(\text{EQ}_{\hat{d},1}^{\text{cs-ts}})$) using approximately smallest

Table 5 The eigenvalue minimization for the nc chained Wood function with $\hat{d} = 2, k = 1$

n	Sparse			Dense		
	mb	opt	Time	mb	opt	Time
20	3	1.0000	0.05	31	1.0000	0.16
40	3	0.9997	0.08	61	1.0000	1.14
60	3	0.9992	0.09	91	1.0000	7.06
80	3	1.0000	0.10	121	1.0000	30.9
100	3	1.0000	0.10	151	1.0000	100
200	3	0.9978	0.16	–	–	–
400	3	0.9930	0.43	–	–	–
600	3	0.9871	0.71	–	–	–
800	3	0.9846	1.04	–	–	–
1000	3	0.9919	1.41	–	–	–
2000	3	0.9605	4.95	–	–	–
3000	3	0.9889	9.93	–	–	–
4000	3	0.9652	18.6	–	–	–

Table 6 The eigenvalue minimization for the nc Broyden tridiagonal function with $\hat{d} = 2, k = 1$

n	Sparse			Dense		
	mb	opt	Time	mb	opt	Time
20	5	0	0.07	41	0	0.25
40	5	0	0.08	81	0	3.28
60	5	0	0.09	121	0	21.6
80	5	0	0.13	161	0	117
100	5	0	0.15	201	0	335
200	5	0	0.29	–	–	–
400	5	0	0.66	–	–	–
600	5	0	0.97	–	–	–
800	5	0	1.56	–	–	–
1000	5	0	2.17	–	–	–
2000	5	0	7.58	–	–	–
3000	5	0	17.0	–	–	–
4000	5	0	29.5	—	–	–

chordal extensions with $\hat{d} = 3$ (the minimum relaxation order). The results are reported in Table 7. To show the benefits of our method by contrast with the usual sparse approach based on correlative sparsity, we also display the results for the latter approach (i.e., $(\mathrm{EQ}^{\mathrm{cs}}_{\hat{d}})$) and the results for the dense approach in the table. Again one can see from the table that our sparse approach is more scalable than the approach that exploits only correlative sparsity as well as the dense approach. Actually, the last two can never be executed due to the memory limit even when the problem has only 6 variables.

Table 7 The eigenvalue minimization for the nc Broyden banded function over $\mathcal{D}$ with $\hat{d} = 3, k = 1$

n	CS+TS			CS			Dense		
	mb	opt	Time	mb	opt	Time	mb	opt	Time
5	11	3.113	0.50	156	3.113	70.7	156	3.113	69.8
10	15	3.011	2.78	400	–	–	–	–	–
20	15	9.658	11.4	400	–	–	–	–	–
30	15	16.30	22.3	400	–	–	–	–	–
40	15	22.94	38.1	400	–	–	–	–	–
50	15	29.57	57.7	400	–	–	–	–	–
60	15	36.21	80.5	400	–	–	–	–	–
70	15	42.85	105	400	–	–	–	–	–
80	15	49.49	138	400	–	–	–	–	–
90	15	56.13	151	400	–	–	–	–	–
100	15	62.77	180	400	–	–	–	–	–

"CS + TS" indicates the results for the approach that exploits combined term-correlative sparsity; "CS" indicates the results for the approach that exploits only correlative sparsity

Table 8 The maximal chordal extension (max) versus approximate smallest chordal extensions (min) with $\hat{d} = 2, k = 1$

t	max			min		
	mb	opt	Time	mb	opt	Time
80	34	−0.5366	0.26	2	−0.5366	0.15
100	37	−0.6289	0.34	3	−0.6289	0.12
120	49	−0.7885	0.48	3	−0.7946	0.19
140	54	−0.6265	0.80	3	−0.6265	0.17
160	88	−0.7464	6.37	3	−0.7486	0.34
180	110	−0.6357	17.1	3	−0.6360	0.23
200	117	−0.6899	24.1	3	−0.6930	0.35
220	141	−0.7878	51.2	3	−0.7901	0.38
240	178	−0.7633	213	4	−0.7780	0.36
260	207	–	–	5	−0.8190	0.48

6.2 Randomly generated examples

To compare the performance of the maximal chordal extension with approximate smallest chordal extensions, let us consider the eigenvalue minimization for a random quartic polynomial f over the unit ball. To this end, let $h \in \mathbb{R}\langle X_1, \ldots, X_{20}\rangle$ be a random quartic nc polynomials with t terms and coefficients taken from $[-1, 1]$, and let $f = (h + h^\star)/2$. The results of solving $(\mathrm{EQ}^{\mathrm{ts}}_{2,1})$ are provided in Table 8. It can be seen that even though approximate smallest chordal extensions might give a slightly

Table 9 The eigenvalue minimization for randomly generated examples over multi-balls with $\hat{d} = 2$

n	CS+TS				CS			dense		
	k	mb	opt	time	mb	opt	time	mb	opt	time
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						

"CS + TS" indicates the results for the approach that exploits combined term-correlative sparsity; "CS" indicates the results for the approach that exploits only correlative sparsity

looser bound, it yields significantly smaller block sizes and hence is more efficient than using the maximal chordal extension.

Now we construct randomly generated examples whose csp graph consists of p maximal cliques of size 15 as follows: let $f = \sum_{l=1}^{p}(h_l + h_l^\star)/2$ where $h_l \in \mathbb{R}\langle X_{10l-9}, \ldots, X_{10l+5}\rangle$ is a random quartic nc polynomials with 15 terms and coefficients taken from $[-1, 1]$, and let $S = \{g_l\}_{l=1}^{p}$ where $g_l = 1 - X_{10l-9}^2 - \cdots - X_{10l+5}^2$. We consider the eigenvalue minimization problem for f over the multi-ball $\mathbb{B}$ defined by S. Let $p = 50, 100, \ldots, 400$ so that we obtain 8 such instances[2]. We compute the NCTSSOS hierarchy $(\lambda_{\hat{d},k}^{\text{cs-ts}}(f, S))_{k\geq 1}$ with $\hat{d} = 2$ and report the results of the first three steps (where we use the maximal chordal

[2] The polynomials are available at https://wangjie212.github.io/jiewang/code.html.

Table 10 The trace minimization for the nc Broyden banded function with $\hat{d} = 3, k = 1$

n	Sparse			Dense		
	mb	opt	time	mb	opt	time
10	29	0	1.91	–	–	–
20	29	0	9.72	–	–	–
30	29	0	18.2	–	–	–
40	29	0	34.3	–	–	–
50	29	0	46.2	–	–	–
60	29	0	65.4	–	–	–
70	29	0	79.5	–	–	–
80	29	0	99.1	–	–	–
90	29	0	118	–	–	–
100	29	0	150	–	–	–

Table 11 The trace minimization for the nc Broyden tridiagonal function with $\hat{d} = 2, k = 1$

n	Sparse			Dense		
	mb	opt	Time	mb	opt	Time
20	6	0	0.16	–	–	–
40	6	0	0.27	–	–	–
60	6	0	0.36	–	–	–
80	6	0	0.44	–	–	-
100	6	0	0.57	–	–	–
200	6	0	1.36	–	–	–
400	6	0	3.48	–	–	–
600	6	0	7.28	–	–	–
800	6	0	10.9	–	–	–
1000	6	0	15.4	–	-	–
2000	6	0	55.9	–	–	–
3000	6	0	122	–	–	–
4000	6	0	220	–	–	–

extension for the first step and use approximate smallest chordal extensions for the second and third steps, respectively) in Table 9. As one may expect, neither the dense approach nor the approach that exploits only correlative sparsity can handle problems with so large sizes. On the other hand, our sparse approach easily scales up to 4005 variables.

6.3 Trace optimization examples

Let us consider the unconstrained trace minimization for the nc Broyden banded function and the nc Broyden tridiagonal function. We compute $\mu_1^{\text{cs-ts}}(f)$ (the optimum of $(\text{EQ}_1^{\text{cs-ts}})$) using approximately smallest chordal extensions. The results are

Table 12 The trace minimization for the nc Broyden banded function over $\mathcal{D}$ with $\hat{d} = 3, k = 1$

n	CS + TS			CS			Dense		
	mb	opt	Time	mb	opt	Time	mb	opt	Time
5	19	3.113	0.66	156	3.113	7.24	156	3.113	7.10
10	29	3.011	5.88	400	–	–	–	–	–
20	29	9.833	32.9	400	–	–	–	–	–
30	29	16.47	49.5	400	–	–	–	–	–
40	29	23.11	73.4	400	–	–	–	–	–
50	29	29.75	111	400	–	–	–	–	–
60	29	36.39	151	400	–	–	–	–	–
70	29	43.03	198	400	–	–	–	–	–
80	29	49.67	238	400	–	–	–	–	–
90	29	56.31	298	400	–	–	–	–	–
100	29	62.95	338	400	–	–	–	–	–

"CS+TS" indicates the results for the approach that exploits combined term-correlative sparsity; "CS" indicates the results for the approach that exploits only correlative sparsity

reported in Table 10–11, respectively. As for eigenvalue minimization, the sparse approach is much more scalable than the dense approach, which actually can never be executed for these examples due to the memory limit.

We next consider the trace minimization for the nc Broyden banded function over the semialgebraic set $\mathcal{D}$ defined in Sect. 6.1. We compute $\mu_{\hat{d},1}^{\text{cs-ts}}(f, S)$ (the optimum of $(\text{TQ}_{\hat{d},1}^{\text{cs-ts}})$) using approximately smallest chordal extensions and compare with the results for the approach that exploits only correlative sparsity and the results for the dense approach. The minimum relaxation order $\hat{d} = 3$ is used. The results are reported in Table 12, which again demonstrate the scalability of our sparse approach.

Finally, we consider the trace minimization for the randomly generated quartic nc polynomials used in Sect. 6.1 over the multi-ball $\mathbb{B}$. We compute the NCTSSOS hierarchy $(\mu_{\hat{d},k}^{\text{cs-ts}}(f, S))_{k\geq 1}$ with the relaxation order $\hat{d} = 2$. We report the results of the first three steps (where we always use approximate smallest chordal extensions) in Table 13. As one could expect, neither the dense approach nor the approach that exploits only correlative sparsity can handle these problems. On the other hand, our sparse approach is easily scalable up to 4005 variables.

7 Conclusions and outlook

We have presented the sparsity (term sparsity and combined correlative-term sparsity) adapted moment-SOHS hierarchies for both eigenvalue optimization and trace optimization involving noncommutative polynomials. Numerical experiments demonstrate that these sparse hierarchies are very efficient and scale well with the problem size when appropriate sparsity patterns are accessible. One question left for

Table 13 The trace minimization for randomly generated examples over multi-balls with $\hat{d} = 2$

n	CS+TS				CS			dense		
	k	mb	opt	time	mb	opt	time	mb	opt	time
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						

"CS + TS" indicates the results for the approach that exploits combined term-correlative sparsity; "CS" indicates the results for the approach that exploits only correlative sparsity

future investigation is to develop a Gelfand-Naimark-Segal's style construction for extracting a minimizer adapted to our sparse settings.

Recently a moment-SOHS hierarchy for optimization problems involving trace polynomials was proposed in [22]. It would be worth extending further our sparsity-exploiting framework to handle trace polynomials.

We also plan to use the sparsity adapted moment-SOHS hierarchies developed in this paper to tackle large-scale NCPOPs arising from quantum information and condensed matter physics.

Acknowledgements Both authors were supported by the Tremplin ERC Stg Grant ANR-18-ERC2-0004-01 (T-COPS project). The second author was supported by the FMJH Program PGMO (EPICS project) and EDF, Thales, Orange et Criteo, as well as the PHC Proteus Program (QUANTPOP project n°46195TA). This work has benefited from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Actions, grant agreement 813211 (POEMA)

as well as from the AI Interdisciplinary Institute ANITI funding, through the French "Investing for the Future PIA3" program under the Grant agreement n°ANR-19-PI3A-0004.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Agler, J., Helton, W., McCullough, S., Rodman, L.: Positive semidefinite matrices with a given sparsity pattern. Linear Algebra Appl. **107**, 101–149 (1988)
2. Anjos, M.F., Lasserre, J.B. (eds.): Handbook on semidefinite, conic and polynomial optimization. International Series in Operations Research & Management Science
3. ApS, M.: The MOSEK optimization toolbox. Version 8.1. (2017). http://docs.mosek.com/8.1/toolbox/index.html
4. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: a fresh approach to numerical computing. SIAM Rev. **59**(1), 65–98 (2017)
5. Blair, J.R., Peyton, B.: An introduction to chordal graphs and clique trees. In: Graph Theory and Sparse Matrix Computation. pp. 1–29. Springer (1993)
6. Bodlaender, H.L., Koster, A.M.: Treewidth computations I: upper bounds. Inf. Comput. **208**(3), 259–275 (2010)
7. Bromberger, S., Fairbanks, J., other contributors: JuliaGraphs/LightGraphs.jl: an optimized graphs package for the julia programming language (2017). 10.5281/zenodo.889971
8. Bugarin, F., Henrion, D., Lasserre, J.B.: Minimizing the sum of many rational functions. Math. Programm. Comput. **8**(1), 83–111 (2016)
9. Burgdorf, S., Cafuta, K., Klep, I., Povh, J.: The tracial moment problem and trace-optimization of polynomials. Math. Program. **137**(1–2), 557–578 (2013). https://doi.org/10.1007/s10107-011-0505-8
10. Burgdorf, S., Cafuta, K., Klep, I., Povh, J.: The tracial moment problem and trace-optimization of polynomials. Math. Program. **137**(1), 557–578 (2013)
11. Burgdorf, S., Klep, I., Povh, J.: Optimization of polynomials in non-commuting variables, vol. 2. Springer, bERLIN (2016)
12. Cafuta, K., Klep, I., Povh, J.: Constrained polynomial optimization problems with noncommuting variables. SIAM J. Optim. **22**(2), 363–383 (2012). https://doi.org/10.1137/110830733
13. Chen, T., Lasserre, J.B., Magron, V., Pauwels, E.: Polynomial optimization for bounding Lipschitz constants of deep networks. arXiv preprint arXiv:2002.03657 (2020)
14. Dunning, I., Huchette, J., Lubin, M.: JuMP: A modeling language for mathematical optimization. SIAM Rev. **59**(2), 295–320 (2017)
15. Gribling, S., De Laat, D., Laurent, M.: Lower bounds on matrix factorization ranks via noncommutative polynomial optimization. Found. Comput. Math. **19**(5), 1013–1070 (2019)
16. Gribling, S., de Laat, D., Laurent, M.: Bounds on entanglement dimensions and quantum graph parameters via noncommutative polynomial optimization. Math. Program. **170**(1), 5–42 (2018)
17. Grone, R., Johnson, C.R., Sá, E.M., Wolkowicz, H.: Positive definite completions of partial Hermitian matrices. Linear Algebra Appl. **58**, 109–124 (1984)
18. Helton, J., McCullough, S.: A Positivstellensatz for non-commutative polynomials. Trans. Am. Math. Soc. **356**(9), 3721–3737 (2004)
19. Helton, J.W.: Positive noncommutative polynomials are sums of squares. Ann. Math. **156**, 675–694 (2002)
20. Josz, C., Molzahn, D.K.: Lasserre hierarchy for large scale polynomial optimization in real and complex variables. SIAM J. Optim. **28**(2), 1017–1048 (2018)
21. Klep, I., Magron, V., Povh, J.: Sparse noncommutative polynomial optimization. Math Programm. (2021). https://doi.org/10.1007/s10107-020-01610-1

22. Klep, I., Magron, V., Volčič, J.: Optimization over trace polynomials. arXiv preprint arXiv:2006.12510 (2020)
23. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**(3), 796–817 (2001)
24. Lasserre, J.B.: Convergent SDP-relaxations in polynomial optimization with sparsity. SIAM J. Optim. **17**(3), 822–843 (2006)
25. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. Emerg. Appl. Algebr. Geom. (2009). https://doi.org/10.1007/978-0-387-09686-5_7
26. Magron, V.: Interval enclosures of upper bounds of roundoff errors using semidefinite programming. ACM Trans. Math. Softw. **44**(4), 1–18 (2018)
27. Magron, V., Constantinides, G., Donaldson, A.: Certified roundoff error bounds using semidefinite programming. ACM Trans. Math. Softw. **43**(4), 1–34 (2017)
28. Mai, N.H.A., Lasserre, J.B., Magron, V., Wang, J.: Exploiting constant trace property in large-scale polynomial optimization. arXiv preprint arXiv:2012.08873 (2020)
29. Mai, N.H.A., Magron, V., Lasserre, J.B.: A sparse version of Reznick's Positivstellensatz. arXiv preprint arXiv:2002.05101 (2020)
30. Marecek, J., Vala, J.: Quantum optimal control via magnus expansion: The non-commutative polynomial optimization problem. arXiv preprint arXiv:2001.06464 (2020)
31. Navascués, M., Pironio, S., Acín, A.: A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. New J. Phys. **10**(7), 073013 (2008)
32. Pál, K.F., Vértesi, T.: Quantum bounds on Bell inequalities. Phys. Rev. A **79**(2), 022120 (2009)
33. Pironio, S., Navascués, M., Acín, A.: Convergent relaxations of polynomial optimization problems with noncommuting variables. SIAM J. Optim. **20**(5), 2157–2180 (2010). https://doi.org/10.1137/090760155
34. Pironio, S., Navascués, M., Acin, A.: Convergent relaxations of polynomial optimization problems with noncommuting variables. SIAM J. Optim. **20**(5), 2157–2180 (2010)
35. Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. **42**(3), 969–984 (1993). https://doi.org/10.1512/iumj.1993.42.42045
36. Reznick, B., et al.: Extremal PSD forms with few terms. Duke Math. Jo. **45**(2), 363–374 (1978)
37. Skelton, R.E., Iwasaki, T., Grigoriadis, D.E.: A Unified Algebraic Approach to Control Design. CRC Press, Boca Raton (1997)
38. Tacchi, M., Weisser, T., Lasserre, J.B., Henrion, D.: Exploiting sparsity for semi-algebraic set volume computation. Foundations of Computational Mathematics pp. 1–49 (2021)
39. Takesaki, M.: Theory of operator algebras. III, *Encyclopaedia of Mathematical Sciences*, vol. 127. Springer-Verlag, Berlin (2003). https://doi.org/10.1007/978-3-662-10453-8. Operator Algebras and Non-commutative Geometry, 8
40. Vandenberghe, L., Andersen, M.S., et al.: Chordal graphs and semidefinite optimization. Found. Trends Regist. Optim. **1**(4), 241–433 (2015)
41. Waki, H., Kim, S., Kojima, M., Muramatsu, M.: Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity. SIAM J. Optim. **17**(1), 218–242 (2006)
42. Waki, H., Kim, S., Kojima, M., Muramatsu, M., Sugimoto, H.: Algorithm 883: SparsePOP: a sparse semidefinite programming relaxation of polynomial optimization problems. ACM Trans. Math. Softw. **35**(2), 1–13 (2008)
43. Wang, J.: ChordalGraph: A Julia package to handle chordal graphs (2020). https://github.com/wangjie212/ChordalGraph
44. Wang, J., Li, H., Xia, B.: A new sparse SOS decomposition algorithm based on term sparsity. In: Proceedings of the 2019 on international symposium on symbolic and algebraic computation, pp. 347–354 (2019)
45. Wang, J., Maggio, M., Magron, V.: SparseJSR: A fast algorithm to compute joint spectral radius via sparse SOS decompositions. In: 2021 American Control Conference. IEEE (2021)
46. Wang, J., Magron, V., Lasserre, J.B.: Chordal-TSSOS: a moment-SOS hierarchy that exploits term sparsity with chordal extension. SIAM J. Optim. **31**(1), 114–141 (2021)
47. Wang, J., Magron, V., Lasserre, J.B.: TSSOS: a moment-SOS hierarchy that exploits term sparsity. SIAM J. Optim. **31**(1), 30–58 (2021)
48. Wang, J., Magron, V., Lasserre, J.B., Mai, N.H.A.: CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization. arXiv:2005.02828 (2020)

49. Yurtsever, A., Tropp, J.A., Fercoq, O., Udell, M., Cevher, V.: Scalable semidefinite programming. SIAM J. Math. Data Sci. **3**(1), 171–200 (2021)
50. Zhou, Q., Marecek, J.: Proper learning of linear dynamical systems as a non-commutative polynomial optimisation problem. arXiv preprint arXiv:2002.01444 (2020)
51. Zhou, Q., Marecek, J., Shorten, R.N.: Fairness in forecasting and learning linear dynamical systems. arXiv preprint arXiv:2006.07315 (2020)