FULL LENGTH PAPER

Series A

Exploiting ideal-sparsity in the generalized moment problem with application to matrix factorization ranks

Milan Korda[1,2] · Monique Laurent[3,4] · Victor Magron[1] · Andries Steenkamp[3]

Abstract

We explore a new type of sparsity for the generalized moment problem (GMP) that we call *ideal-sparsity*. In this setting, one optimizes over a measure restricted to be supported on the variety of an ideal generated by quadratic bilinear monomials. We show that this restriction enables an equivalent sparse reformulation of the GMP, where the single (high dimensional) measure variable is replaced by several (lower dimensional) measure variables supported on the maximal cliques of the graph corresponding to the quadratic bilinear constraints. We explore the resulting hierarchies of moment-based relaxations for the original dense formulation of GMP and this new, equivalent ideal-sparse reformulation, when applied to the problem of bounding nonnegative- and completely positive matrix factorization ranks. We show that the ideal-sparse hierar-

✉ Monique Laurent
   monique.laurent@cwi.nl

   Milan Korda
   milan.korda@laas.fr

   Victor Magron
   vmagron@laas.fr

   Andries Steenkamp
   andries.steenkamp@cwi.nl

[1] LAAS-CNRS and Institute of Mathematics, Toulouse, France

[2] Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic

[3] Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands

[4] Tilburg University, Tilburg, The Netherlands

Published online: 08 July 2023   ⌀ Springer

chies provide bounds that are at least as good (and often tighter) as those obtained from the dense hierarchy. This is in sharp contrast to the situation when exploiting correlative sparsity, as is most common in the literature, where the resulting bounds are weaker than the dense bounds. Moreover, while correlative sparsity requires the underlying graph to be chordal, no such assumption is needed for ideal-sparsity. Numerical results show that the ideal-sparse bounds are often tighter and much faster to compute than their dense analogs.

Keywords Generalized moment problem · Polynomial optimization · Sparsity · Matrix factorization rank · Completely positive rank · Nonnegative rank · Semidefinite programming

Mathematics Subject Classification 15A23 · 81P40 · 81P42 · 90C22 · 90C23

1 Introduction

We consider the *generalized moment problem* (abbreviated as GMP), of the form

$$\text{val} := \inf_{\mu \in \mathcal{M}(\mathbb{R}^n)} \left\{ \int f_0 d\mu : \int f_i d\mu = a_i \ (i \in [N]), \ \text{Supp}(\mu) \subseteq K \right\}, \quad (1)$$

where $f_0, f_i \in \mathbb{R}[x]$ are multivariate polynomials in the variables $x = (x_1, ..., x_n)$, $a_i \in \mathbb{R}$, $K \subseteq \mathbb{R}^n$ (taken to be Borel measurable), and the optimization is over the set $\mathcal{M}(\mathbb{R}^n)$ of (finite positive) Borel measures on $\mathbb{R}^n$. In problem (1) we restrict to measures $\mu \in \mathcal{M}(\mathbb{R}^n)$ whose support $\text{Supp}(\mu)$ is contained in K, which is equivalent to requiring $\int f d\mu = \int_K f d\mu$ for any Borel measurable function $f : \mathbb{R}^n \to \mathbb{R}$. Throughout, we assume that K is a basic closed semialgebraic set of the form

$$K = \{x \in \mathbb{R}^n : g_j(x) \geq 0 \ (j \in [m]), \ x_i x_j = 0 \ (\{i, j\} \in \overline{E})\}, \quad (2)$$

where $g_j \in \mathbb{R}[x]$ are polynomials, E is a given set of pairs of distinct elements of $V = [n] := \{1, \ldots, n\}$, and $\overline{E}$ is the following set of pairs

$$\overline{E} = \{\{i, j\} : i \in V, j \in V, i \neq j, \{i, j\} \notin E\}.$$

Hence, the set K is contained in the variety of the ideal

$$I_E := \left\{ \sum_{\{i,j\} \in \overline{E}} u_{ij} x_i x_j : u_{ij} \in \mathbb{R}[x] \right\} \subseteq \mathbb{R}[x] \quad (3)$$

generated by the monomials $x_i x_j$ for the pairs $\{i, j\} \in \overline{E}$. It will be convenient to consider the graph $G = (V, E)$, so that the conditions $x_i x_j = 0$ appearing in the definition of K correspond to the *nonedges* of G. This notation may seem at first sight cumbersome. However, the motivation for it is that the graph G functions as supporting solutions for problem (1); this will be especially useful for applications to matrix

factorization ranks like the completely positive rank (cp-rank) or the nonnegative rank of a matrix, where G will correspond to the support graph of the matrix.

The generalized moment problem (with K semialgebraic) has been much studied in recent years. It permits to model a wide variety of problems, including polynomial optimization (minimization of a polynomial or rational function over K), volume computation, control theory, option pricing in finance, and much more. See, e.g., [36, 45–47] and further references therein.

The focus of this paper is to exploit the presence of explicit ideal constraints (of a special form) in the description of the semialgebraic set K for solving problem (1). This indeed naturally implies some sparsity structure on problem (1), to which we will refer as *ideal-sparsity* structure. Our objective is to explore how one can best exploit this ideal-sparsity structure in order to define more efficient semidefinite hierarchies for problem (1) and apply them to sparse matrix factorization ranks. A remarkable feature is that the ideal-sparse hierarchies provide bounds that are at least as good (and often better) as the bounds provided by the original dense hierarchy. Moreover, the underlying sparsity graph is not required to be chordal. Both these features are in stark contrast to the existing sparse hierarchies based on correlative sparsity whose bounds are always dominated by the dense bounds and that require the underlying sparsity graph to be chordal in order to guarantee convergence. We refer to Sect. 3.2 for an in-depth discussion about correlative and ideal-sparsity.

We focus here on the application to the completely positive and the nonnegative factorization ranks, asking for a factorization by nonnegative vectors. However, as we will mention in the final discussion section, this ideal-sparsity framework could also be applied to more general settings. Indeed, it could be applied to other matrix factorization ranks, such as the (completely) positive semidefinite rank, where one asks for a factorization by positive semidefinite matrices, in which case one would have to apply tools from polynomial optimization in noncommutative variables. Also, instead of an ideal generated by quadratic monomials, one could have an ideal generated by higher degree monomials. In addition, up to a change of variables, one could consider an ideal generated by more general products of linear terms, such as $(a^T x + b)(c^T x + d)$. This type of constraint, often known as a *complementary constraint*, occurs in various applications, including ReLU neural networks or optimization when considering KKT optimality conditions.

Next, we mention the overall organization of the paper and give some general notation used throughout. After that, we will give a broad overview of the contents and main results obtained in the paper.

1.1 Organization of the paper

The paper is organized as follows. In the rest of the Introduction we outline the main results in the paper. Then, in Sect. 2 we recall some preliminaries about linear functionals on polynomials and moment matrices. In Sect. 3 we consider the GMP (1): we show its sparse reformulation (11), we present the corresponding sparse hierarchies, and we discuss how ideal-sparsity relates to the more classic correlative sparsity. Sec-

tion 4 is devoted to the application to the cp-rank and Sect. 5 to the application to the nonnegative rank. We conclude with some final remarks and discussions in Sect. 6.

1.2 Notation

We gather here some notation that is used throughout the paper. For $n, t \in \mathbb{N}$ set $\mathbb{N}_t^n = \{\alpha \in \mathbb{N}^n : |\alpha| \leq t\}$, where $|\alpha| = \sum_{i=1}^n \alpha_i$ denotes the degree of the monomial $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. We let $[x]_t = (x^\alpha)_{\alpha \in \mathbb{N}_t^n}$ denote the vector of monomials with degree at most t (listed in some given fixed order). Moreover, $\mathbb{R}[x]$ (resp., $\mathbb{R}[x]_t$) denotes the set of n-variate polynomials in variables $x = (x_1, \ldots, x_n)$ (with degree at most t). Let Σ denote the set of sum-of-squares polynomials, of the form $\sum_i q_i^2$ for some $q_i \in \mathbb{R}[x]$, and set $\Sigma_t = \Sigma \cap \mathbb{R}[x]_t$.

Consider a set $U \subseteq [n]$. Given a vector $y \in \mathbb{R}^{|U|}$, we let $(y, 0_{V \setminus U}) \in \mathbb{R}^n$ denote the vector obtained by padding y with zeros at the entries indexed by $[n] \setminus U$. For an n-variate function $f : \mathbb{R}^{|V|} \to \mathbb{R}$, we let $f_{|U} : \mathbb{R}^{|U|} \to \mathbb{R}$ denote the function in the variables $x(U) = \{x_i : i \in U\}$, which is obtained from f by setting to zero all the variables x_i indexed by $i \in V \setminus U$. That is, $f_{|U}(y) = f(y, 0_{V \setminus U})$ for $y \in \mathbb{R}^{|U|}$. So, if f is an n-variate polynomial, then $f_{|U}$ is a $|U|$-variate polynomial in the variables $x(U)$.

For a symmetric matrix $M \in \mathcal{S}^n$, the notation $M \succeq 0$ means that M is positive semidefinite, i.e., $v^T M v \geq 0$ for all $v \in \mathbb{R}^n$. Throughout, we let I_n and J_n denote the identity matrix and the all-ones matrix of size n, which we sometimes also denote as I and J when the dimension is clear from the context. The support of a vector $x \in \mathbb{R}^n$ is the set $\mathrm{Supp}(x) = \{i \in [n] : x_i \neq 0\}$.

1.3 Roadmap through the paper

In the rest of this section, we now offer a quick roadmap through the main contents of the paper. We begin with recalling how to define the dense hierarchy of bounds for the problem (1). We then discuss their main drawback (quick growth of the matrices involved in the semidefinite programs) and several options for addressing this difficulty that have been offered in the literature. After that, we introduce the new ideal-sparse reformulation of problem (1) and the corresponding ideal-sparse hierarchy, which we then specialize to the applications for bounding the completely positive and nonnegative ranks.

1.3.1 Classical (dense) moment relaxations

We begin by recalling the classical moment approach that permits to build hierarchies of semidefinite approximations for problem (1). For details, see, e.g., the monograph by Lasserre [46], or the survey [21]. For $t \in \mathbb{N} \cup \{\infty\}$, the set

$$\mathcal{M}(\mathbf{g})_{2t} = \left\{ \sum_{j=0}^m \sigma_j g_j : \sigma_j \in \Sigma, \ \deg(\sigma_j g_j) \leq 2t \right\} \subseteq \mathbb{R}[x]_{2t} \qquad (4)$$

is the quadratic module generated by $\mathbf{g} = (g_1, \ldots, g_m)$, and truncated at degree $2t$, setting $g_0 = 1$. We also set $\mathcal{M}(\mathbf{g}) = \mathcal{M}(\mathbf{g})_\infty$. Here, Σ denotes the set of sums of squares of polynomials in $\mathbb{R}[x]$. Similarly,

$$I_{E,2t} = \left\{ \sum_{\{i,j\} \in \overline{E}} u_{ij} x_i x_j : u_{ij} \in \mathbb{R}[x]_{2t-2} \right\} \subseteq \mathbb{R}[x]_{2t} \tag{5}$$

denotes the truncation of the ideal I_E at degree $2t$. We can now define the moment relaxation of level t for problem (1):

$$\begin{aligned}
\xi_t := \inf\{L(f_0) : \ &L \in \mathbb{R}[x]_{2t}^*, \\
&L(f_i) = a_i \ (i \in [N]), \\
&L \geq 0 \text{ on } \mathcal{M}(\mathbf{g})_{2t}, \ L = 0 \text{ on } I_{E,2t}\}.
\end{aligned} \tag{6}$$

Here, $\mathbb{R}[x]_{2t}^*$ denotes the set of linear functionals $L : \mathbb{R}[x]_{2t} \to \mathbb{R}$. The motivation for the above parameter is as follows. Assume $\mu \in \mathcal{M}(\mathbb{R}^n)$ is a measure that is feasible for problem (1), and consider the associated linear functional L that acts on $\mathbb{R}[x]_{2t}$ via integration: $p \in \mathbb{R}[x]_{2t} \mapsto L(p) = \int p \, d\mu$. Then, it is easy to see that L is feasible for (6): $L(f_i) = \int f_i \, d\mu = a_i$, $L \geq 0$ on $\mathcal{M}(\mathbf{g})_{2t}$ (since any polynomial in $\mathcal{M}(\mathbf{g})_{2t}$ is nonnegative on the set K), and $L = 0$ on $I_{E,2t}$ (since any polynomial in $I_{E,2t}$ vanishes on K). This shows that the parameter ξ_t lower bounds the optimum value val of problem (1).

We refer to the above hierarchy of parameters ξ_t as the *dense moment hierarchy*. Clearly, they satisfy $\xi_t \leq \xi_{t+1} \leq \xi_\infty \leq$ val. Moreover, under some mild assumptions, these bounds converge asymptotically to the optimum value val of (1). This fundamental property follows from the general theory about GMP (see, e.g., [21, 46]) and is summarized in the following theorem that will be used repeatedly throughout this work.

Theorem 1 *Assume problem* (1) *is feasible and the following Slater-type condition holds:*

$$\text{there exist scalars } z_0, z_1, \ldots, z_N \in \mathbb{R} \text{ such that } \sum_{i=0}^N z_i f_i(x) > 0 \text{ for all } x \in K. \tag{7}$$

Then, problem (1) *has an optimal solution* μ, *which can be chosen to be finite atomic. If, in addition,* $\mathcal{M}(\mathbf{g})$ *is Archimedean, i.e.,* $R - \sum_{i=1}^n x_i^2 \in \mathcal{M}(\mathbf{g})$ *for some scalar* $R > 0$, *then we have* $\lim_{t \to \infty} \xi_t = \xi_\infty =$ val.

As it will be recalled in Sect. 2.1, program (6) can be reformulated as a semidefinite program and thus the bound ξ_t can be computed using semidefinite optimization algorithms. However, a common drawback of the dense hierarchy (6) is that it involves matrices whose size grows very quickly with the level t and with the degree and number of variables of the polynomials $f_0, f_1, \ldots, f_N, g_1, \ldots, g_m$. Hence, even though these relaxations are convex, they might be challenging to solve already for GMP instances of modest size.

1.3.2 Existing schemes to improve scalability of the dense moment relaxations

Several schemes have been developed to overcome the scalability issue of the dense hierarchy (6) just mentioned above. They aim to reduce the size of the involved matrices by exploiting the specific structure of the input polynomials without compromising the convergence guarantees of the structure-induced moment relaxations. One workaround consists of exploiting the symmetries [56], but this requires that each input polynomial is invariant under the action of a subgroup of the general linear group.

Another approach is to exploit different kinds of sparsity structures. The first kind is called *correlative sparsity*, which occurs when there are few correlations between the variables of the input polynomials [44, 63]. Correlative sparsity has been extended to derive moment relaxations of polynomial problems in complex variables [39], non-commutative variables [40] and polynomial matrix inequalities [69]. The second kind is called *term sparsity*, which occurs when they are few (by comparison with all possible) monomial terms involved in the input polynomials, and for which correlative sparsity is not exploitable. For unconstrained polynomial optimization, one well-known solution is to eliminate the monomial terms which never appear among the support of sums of squares decompositions [55]. Alternatively, one can decompose the input polynomial as a sum of nonnegative circuits, by solving a geometric programming relaxation [38] or a second-order cone programming relaxation [5, 64], or as a sum of arithmetic–geometric-mean-exponentials [15] with relative entropy programming relaxations. Term sparsity has recently been the focus of active research with extensions to constrained polynomial optimization [65, 66]. Note that both kinds of sparsity can be combined [67]. For a general exposition about sparse polynomial optimization, we refer to the recent surveys [51, 70].

We will return to the correlative sparsity approach for GMP later in Sect. 3.2 and discuss how it relates to the new ideal-sparsity structure considered in the paper. By contrast with classical polynomial optimization problems, it is not completely clear which initial set of monomials should be chosen to initialize the term sparsity hierarchy when facing a given GMP instance. Therefore, we do not explore the combination of term and ideal-sparsity, for such an investigation would warrant a separate publication.

1.3.3 New ideal-sparse moment relaxations

As we now see, one can exploit the fact that the set K in (2) is contained in the variety of the ideal I_E from (3). The basic idea is that, instead of optimizing over a *single* measure μ supported on $K \subseteq \mathbb{R}^n$, one may optimize over *several* measures that are supported on smaller dimensional spaces.

A set $W \subseteq V$ is a clique of the graph $G = (V, E)$ if $\{u, v\} \in E$ for any two distinct vertices $u, v \in W$. A clique is maximal (w.r.t inclusion) if it is not strictly contained in any other clique of G. Let $V_1, \ldots, V_p$ denote the maximal cliques of the graph $G = (V, E)$ and, for $k \in [p]$, define the following subset of K:

$$\widehat{K_k} := \{x \in K : \mathrm{Supp}(x) \subseteq V_k\} \subseteq K \subseteq \mathbb{R}^n. \tag{8}$$

Recall $\text{Supp}(x) = \{i \in [n] : x_i \neq 0\}$ denotes the support of $x \in \mathbb{R}^n$. If $x \in K$, then its support $\text{Supp}(x)$ is a clique of the graph G and thus it is contained in a maximal clique V_k, so that $x \in \widehat{K}_k$ for some $k \in [p]$. Therefore, the sets $\widehat{K}_1, \ldots, \widehat{K}_p$ cover the set K:

$$K = \widehat{K}_1 \cup \ldots \cup \widehat{K}_p. \tag{9}$$

Next, define the projection $K_k \subseteq \mathbb{R}^{|V_k|}$ of $\widehat{K}_k$ onto the subspace indexed by V_k:

$$K_k := \{y \in \mathbb{R}^{|V_k|} : (y, 0_{V \setminus V_k}) \in \widehat{K}_k\} \subseteq \mathbb{R}^{|V_k|}. \tag{10}$$

Recall that $(y, 0_{V \setminus V_k})$ denotes the vector of $\mathbb{R}^n$ obtained from $y \in \mathbb{R}^{|V_k|}$ by padding it with zeros at all entries indexed by $V \setminus V_k$. Moreover, given a function $f : \mathbb{R}^{|V|} \to \mathbb{R}$, the function $f_{|V_k}(y) : \mathbb{R}^{|V_k|} \to \mathbb{R}$ is defined by $f_{|V_k}(y) = f(y, 0_{V \setminus V_k})$ for $y \in \mathbb{R}^{|V_k|}$. We may now define the following sparse analog of problem (1):

$$\text{val}^{\text{isp}} := \inf_{\mu_k \in \mathcal{M}(\mathbb{R}^{|V_k|}), k \in [p]} \left\{ \sum_{k=1}^{p} \int f_{0|V_k} d\mu_k : \sum_{k=1}^{p} \int f_{i|V_k} d\mu_k = a_i \ (i \in [N]), \right.$$
$$\left. \text{Supp}(\mu_k) \subseteq K_k \ (k \in [p]) \right\}. \tag{11}$$

Hence, while problem (1) has a single measure variable μ on the space $\mathbb{R}^{|V|}$, problem (11) involves p measure variables, where μ_k is on the smaller dimensional space $\mathbb{R}^{|V_k|}$. As we will show in Proposition 6 below, both formulations (1) and (11) are in fact equivalent, i.e, we have equality $\text{val} = \text{val}^{\text{isp}}$. Here, we use the superscript 'isp' as a reminder that the formulation exploits ideal-sparsity; we will follow this same notation below for the corresponding moment hierarchy and also later for the parameters attached to matrix factorization ranks.

Based on its reformulation via (11), we can now define another hierarchy of moment approximations for problem (1), to which we refer as the *ideal-sparse moment hierarchy*:

$$\xi_t^{\text{isp}} := \inf \left\{ \sum_{k=1}^{p} L_k(f_{0|V_k}) : L_k \in \mathbb{R}[x(V_k)]_{2t}^* \ (k \in [p]), \right.$$
$$\sum_{k=1}^{p} L_k(f_{i|V_k}) = a_i \ (i \in [N]), \tag{12}$$
$$\left. L_k \geq 0 \text{ on } \mathcal{M}(\mathbf{g}_{|V_k})_{2t} \ (k \in [p]) \right\}.$$

This hierarchy provides bounds for val that are at least at good as the bounds (6). Namely,

$$\xi_t \leq \xi_t^{\text{isp}} \leq \text{val}$$

holds for any $t \geq 1$ (see Theorem 7 below).

Hence, the ideal-sparse bounds ξ_t^{isp} present a double advantage compared to the dense bounds ξ_t. First, they are at least as good and sometimes strictly better, as we will see later in concrete examples. For the application to the completely positive and nonnegative ranks, we will see classes of matrices showing a large separation between the dense bound and the ideal-sparse bound of level $t = 1$; see Examples 14 and 16. Second, their computation is potentially faster since the sets V_k can be much smaller than the full set V. We will also see in later examples that the computation of the ideal-sparse bounds can be much faster indeed. On the other hand, the number of cliques in the graph G could be large, so there is a trade-off. We refer to discussions later in the paper around specific applications.

Interestingly, no structural chordality property needs to be assumed on the cliques $V_1, \ldots, V_p$ of the graph G. We will comment in Sect. 3.2 about the link between the ideal-sparsity approach presented here and the more classical correlative sparsity approach that can be followed when considering a chordal extension $\widehat{G}$ of the graph G.

The idea of optimizing over multiple measures has appeared already in several contexts, similarly to what can be routinely done in most computational methods, e.g., finite elements. In the context of analyzing dynamical systems involving polynomial data, a similar trick has been used to perform optimal control of piecewise-affine systems in [1], then later on to characterize invariant measures for piecewise-polynomial systems (see [50, § 3.5]). In the context of set estimation, one can also rely on a multi-measure approach to approximate the moments of Lebesgue measures supported on unions of basic semialgebraic sets [48]. The common idea consists in using the piecewise structure of the dynamics and/or the state-space partition to decompose the measure of interest into a sum of local measures supported on each partition cell. The advantage in our current setting is that these measures are supported on smaller dimensional spaces, which leads to potentially strong computational benefit when considering the associated semidefinite programming relaxations.

We next present instances of GMP to which the above ideal-sparsity framework naturally applies, namely to derive bounds on matrix factorization ranks such as the completely positive rank and the nonnegative rank.

1.3.4 Bounds on the completely positive rank via GMP

Let $A \in \mathcal{S}^n$ be a symmetric matrix with nonnegative entries. Assume A is a completely positive matrix (abbreviated as *cp-matrix*), i.e., A can be written as

$$A = \sum_{\ell=1}^{r} a_\ell a_\ell^T \text{ for some nonnegative vectors } a_1, \ldots, a_r \in \mathbb{R}_+^n.$$

Then, the smallest integer $r \in \mathbb{N}$ for which such a decomposition exists is the *cp-rank* of A, denoted $\text{rank}_{\text{cp}}(A)$. Checking whether a given matrix A is completely positive is a computational hard problem (see [24]). The moment approach has been applied to the question of testing whether A is cp-matrix and finding a cp-factorization, in

particular, by Nie [53] who formulates it as testing the existence of a representing measure (over the standard simplex) for the sequence of entries of A.

Hierarchies of moment-based relaxations have also been employed to obtain sequences of bounds for the rank of tensors [61], as well as for the symmetric nuclear norm of tensors [54]. Here, we focus on the question of bounding the cp-rank. No efficient algorithms are known for finding the cp-rank. This motivates the search for efficient methods giving lower bounds on the cp-rank, as, e.g., in [29, 33, 34]. The following parameter was introduced in [29], as a natural "convexification" of the cp-rank:

$$\tau_{\mathrm{cp}}(A) = \inf \left\{ \lambda : \frac{1}{\lambda} A \in \mathrm{conv}\{ xx^T : x \in \mathbb{R}^n_+, \ A - xx^T \succeq 0, \ A \geq xx^T \} \right\}, \quad (13)$$

providing a lower bound for it: $\tau_{\mathrm{cp}}(A) \leq \mathrm{rank}_{\mathrm{cp}}(A)$. As observed below, the parameter $\tau_{\mathrm{cp}}(A)$ can be reformulated as an instance of problem (1), with an ideal-sparsity structure inherited from the matrix A.

To avoid trivialities we assume $A_{ii} > 0$ for all $i \in [n]$. (Indeed, if A is a cp-matrix with $A_{ii} = 0$, then its i-th row/column is identically zero and thus it can be removed without changing the cp-rank.) Note that the constraints $A \geq xx^T$ and $x \geq 0$ are equivalent[1] to $\sqrt{A_{ii}} x_i - x_i^2 \geq 0$ ($i \in [n]$) and $A_{ij} - x_i x_j \geq 0$ ($1 \leq i < j \leq n$). Moreover, they imply $x_i x_j = 0$ whenever $A_{ij} = 0$. Let us define the graph $G_A = (V, E_A)$ as the *support graph* of A, with

$$E_A = \{\{i, j\} : A_{ij} \neq 0, \ i, j \in V, \ i \neq j\}, \ \overline{E}_A = \{\{i, j\} : A_{ij} = 0, \ i, j \in V, \ i \neq j\}, \quad (14)$$

and define the semialgebraic set

$$K_A = \{ x \in \mathbb{R}^n : \ \sqrt{A_{ii}} x_i - x_i^2 \geq 0 \ (i \in [n]), \ A_{ij} - x_i x_j \geq 0 \ (\{i, j\} \in E_A),$$
$$x_i x_j = 0 \ (\{i, j\} \in \overline{E}_A), \qquad A - xx^T \succeq 0 \}. \quad (15)$$

As we now observe, the parameter $\tau_{\mathrm{cp}}(A)$ can be reformulated as an instance of GMP.

Lemma 2 *The parameter $\tau_{\mathrm{cp}}(A)$ is equal to the optimal value of the generalized moment problem:*

$$\mathrm{val}_{\mathrm{cp}} := \inf_{\mu \in \mathcal{M}(\mathbb{R}^n)} \left\{ \int 1 d\mu : \int x_i x_j d\mu = A_{ij} \ (i, j \in V), \ Supp(\mu) \subseteq K_A \right\}.$$

Proof The (easy) key observation is that any feasible solution to $\tau_{\mathrm{cp}}(A)$, i.e., any decomposition of the form $A = \lambda \sum_{\ell=1}^s \lambda_\ell a_\ell a_\ell^T$ with $\lambda_\ell \geq 0$, $\sum_{\ell=1}^s \lambda_\ell = 1$ and $a_\ell \in K_A$, corresponds to a measure $\mu := \lambda \sum_{\ell=1}^s \lambda_\ell \delta_{a_\ell}$ that is feasible for $\tau_{\mathrm{cp}}(A)$ and finite atomic (and vice-versa). Observe also that the Slater-type condition (7) holds (since $f_0 = 1 > 0$ on K_A). The result now follows using (the first part of) Theorem 1:

---

[1] The reason for using the constraint $\sqrt{A_{ii}} x_i - x_i^2 \geq 0$ instead of $A_{ii} - x_i^2 \geq 0$ is because it leads to a larger quadratic module and thus a possibly better bound (see [33]).

if A is completely positive, then $\mathrm{val}_{\mathrm{cp}}$ is feasible and thus has a finite atomic optimal solution, which implies $\tau_{\mathrm{cp}}(A) = \mathrm{val}_{\mathrm{cp}}$; otherwise, both parameters $\tau_{\mathrm{cp}}(A)$ and $\mathrm{val}_{\mathrm{cp}}$ are infeasible and thus equal to ∞. □

Based on the formulation of the parameter $\tau_{\mathrm{cp}}(A)$ in Lemma 2 as a GMP instance, we can define the corresponding bounds $\xi_t^{\mathrm{cp}}(A)$, obtained as special instance of the bounds (6) (see relations (29)–(34) below). Then, the convergence of the bounds $\xi_t^{\mathrm{cp}}(A)$ to $\tau_{\mathrm{cp}}(A)$ follows as a direct application of Theorem 1.

As in the general case of GMP, one may exploit the presence of the ideal constraints $x_i x_j = 0$ (for $\{i, j\} \in \overline{E}_A$) in the definition of K_A and define a hierarchy of ideal-sparse bounds $\xi_t^{\mathrm{cp,isp}}(A)$. These bounds satisfy

$$\xi_t^{\mathrm{cp}}(A) \leq \xi_t^{\mathrm{cp,isp}}(A) \leq \tau_{\mathrm{cp}}(A) \ \text{ for any } t \geq 1,$$

also with asymptotic convergence to $\tau_{\mathrm{cp}}(A)$. We refer to Sect. 4 for details about these parameters and links to earlier bounds in the literature.

1.3.5 Bounds on the nonnegative rank via GMP

The above approach for the cp-rank naturally extends to the asymmetric setting of the nonnegative rank. For a nonnegative matrix $M \in \mathbb{R}^{m \times n}$, its *nonnegative rank*, denoted $\mathrm{rank}_+(M)$, is defined as the smallest integer r for which there exist nonnegative vectors $a_\ell \in \mathbb{R}_+^m$ and $b_\ell \in \mathbb{R}_+^n$ such that

$$M = \sum_{\ell=1}^r a_\ell b_\ell^T. \tag{16}$$

In other words, $\mathrm{rank}_+(M)$ can be seen as the smallest cp-rank of a cp-matrix $A \in \mathcal{S}^{m+n}$ of the form

$$A = \begin{pmatrix} X & M \\ M^T & Y \end{pmatrix} \ \text{ for some nonnegative symmetric matrices } X \in \mathcal{S}^m, Y \in \mathcal{S}^n.$$

Computing the nonnegative rank is an NP-hard problem [62]. In analogy to the parameter τ_{cp} in (13), the following "convexification" of the nonnegative rank was introduced in [29]:

$$\tau_+(M) = \inf \left\{ \lambda : \frac{1}{\lambda} M \in \mathrm{conv}\{xy^T : x \in \mathbb{R}_+^m, \ y \in \mathbb{R}_+^n, \ M \geq xy^T\} \right\}. \tag{17}$$

Note that, compared to the parameter $\tau_{\mathrm{cp}}(A)$ in (13), where we had an additional constraint $A - xx^T \succeq 0$, we now cannot impose such a constraint.

One can define analogs of the bounds ξ_t^{cp} and $\xi_t^{\mathrm{cp,isp}}$ for the nonnegative rank, which now involve a linear functional acting on polynomials in $m + n$ variables. For convenience, we set $V = [m + n] = U \cup W$, where $U = [m] = \{1, \ldots, m\}$ (corresponding to the row indices of M) and $W = \{m+1, \ldots, m+n\}$ (corresponding

to the column indices of M, up to a shift by m). We also set

$$
\begin{aligned}
E^M &= \{\{i, j\} \in U \times W : M_{i, j-m} \neq 0\}, \\
\overline{E}^M &= (U \times W) \setminus E^M = \{\{i, j\} \in U \times W : M_{i, j-m} = 0\},
\end{aligned} \tag{18}
$$

so that E^M corresponds to the (bipartite) support graph of the matrix M. Note that, in comparison to (14), we now only consider *bipartite* pairs $\{i, j\}$ (with $i \in U$ and $j \in W$). To emphasize the difference between the two situations we now put M as a superscript, while we placed A as subscript in the notation E_A.

Let $M_{\max} = \max_{i,j} M_{ij}$ denote the largest entry of M. As observed in [33], one may assume without loss of generality that the vectors in (16) satisfy $\|a_\ell\|_\infty, \|b_\ell\|_\infty \leq \sqrt{M_{\max}}$ (after rescaling). This motivates defining the following semialgebraic set

$$
K^M = \{x \in \mathbb{R}^{m+n} : \sqrt{M_{\max}} x_i - x_i^2 \geq 0 \ (i \in [m+n]), \ M_{i, j-m} - x_i x_j \geq 0 \ (\{i, j\} \in E^M), \\
x_i x_j = 0 \ (\{i, j\} \in \overline{E}_M\}. \tag{19}
$$

The analog of Lemma 2 holds, which provides a GMP reformulation for $\tau_+(M)$.

Lemma 3 *The parameter $\tau_+(M)$ is equal to the optimal value of the generalized moment problem:*

$$
\inf_{\mu \in \mathcal{M}(\mathbb{R}^{m+n})} \left\{ \int 1 d\mu : \int x_i x_j d\mu = M_{i, j-m} \ (i \in U, j \in W), \ Supp(\mu) \subseteq K^M \right\}.
$$

Based on this formulation of the parameter $\tau_+(M)$, we may consider the corresponding bounds $\xi_t^+(A)$, as special instance of the bounds in (6). Their asymptotic convergence to $\tau_+(A)$ follows as a direct application of Theorem 1. One may also exploit the presence of the ideal constraints $x_i x_j = 0$ (for $\{i, j\} \in \overline{E}^M$) in the definition of K^M and define a hierarchy of sparse bounds $\xi_t^{+,\mathrm{isp}}(M)$. These parameters satisfy

$$
\xi_t^+(M) \leq \xi_t^{+,\mathrm{isp}}(M) \leq \tau_+(M) \ \text{ for any } t \geq 1,
$$

with asymptotic convergence of all parameters to $\tau_+(M)$. We refer to Sect. 5 for details about these parameters.

2 Preliminaries about sums of squares and moments

In this section, we recall some preliminaries about sums of squares and linear functionals on polynomials that we will use throughout. These results are well-known in the polynomial optimization community, we refer, e.g., to the following sources [21, 36, 43, 45–47, 49] and further refereces therein for background and broad overviews.

2.1 Nonnegative linear functionals and moment matrices

The program (6) defining the parameter ξ_t involves a linear functional $L \in \mathbb{R}[x]_{2t}^*$, which is assumed to be nonnegative on the truncated quadratic module $\mathcal{M}(\mathbf{g})_{2t}$ (in (4)) and to vanish on the truncated ideal $I_{E,2t}$ (in (5)). We now recall how these conditions can be expressed more concretely in terms of positive semidefiniteness conditions on associated (moment) matrices and thus used to reformulate the program (6) as a semidefinite program.

For this, given $L \in \mathbb{R}[x]_{2t}^*$, define the matrix

$$M_t(L) := (L(x^\alpha x^\beta))_{\alpha,\beta \in \mathbb{N}_t^n} = L([x]_t[x]_t^T),$$

often called a *(pseudo)moment matrix* in the literature. So, in the notation $L([x]_t[x]_t^T)$, it is understood that L is acting entry-wise on the entries of the polynomial matrix $[x]_t[x]_t^T = (x^{\alpha+\beta})_{\alpha,\beta \in \mathbb{N}_t^n}$. Then, it is well-known (and easy to see) that $L(\sigma) \geq 0$ for all $\sigma \in \Sigma \cap \mathbb{R}[x]_{2t}$ if and only if the matrix $M_t(L)$ is positive semidefinite. Consider now a polynomial g with degree $k = \deg(g)$. Then $L(\sigma g) \geq 0$ for all $\sigma \in \Sigma$ with $\deg(\sigma g) \leq 2t$ if and only if the matrix $M_{t-\lceil k/2 \rceil}(gL) := L(g[x]_{t-\lceil k/2 \rceil}[x]_{t-\lceil k/2 \rceil}^T)$ (often called a *localizing moment matrix*) is positive semidefinite. Hence, the condition $L \geq 0$ on $\mathcal{M}(\mathbf{g})_{2t}$ can be equivalently reformulated via the positive semidefiniteness constraints

$$L([x]_t[x]_t^T) \succeq 0, \quad L(g_j[x]_{t-\lceil \deg(g_j)/2 \rceil}[x]_{t-\lceil \deg(g_j)/2 \rceil}^T) \succeq 0 \text{ for } j \in [m].$$

In the same way, the ideal condition $L = 0$ on $I_{E,2t}$ is equivalent to the linear constraints

$$L(x_i x_j x^\alpha) = 0 \text{ for all } \{i, j\} \in \overline{E} \text{ and } \alpha \in \mathbb{N}_{2t-2}^n.$$

Hence, the parameter ξ_t is expressed as the optimum value of a semidefinite program. Recall that there exist efficient algorithms for solving semidefinite programs up to any precision (under some mild assumptions; see, e.g., [23] and further references therein).

2.2 Flatness and extraction of optimal solutions

As recalled in Theorem 1, if the quadratic module $\mathcal{M}(\mathbf{g})$ is Archimedean (i.e., $R - \sum_i x_i^2 \in \mathcal{M}(\mathbf{g})$ for some $R > 0$), then the bounds ξ_t converge asymptotically to ξ_∞. In addition, if the Slater-type condition (7) holds, then $\xi_\infty = \text{val}$ and problem (1) has a finite atomic optimal solution μ, i.e., supported on finitely many points in K.

A remarkable property of the bounds ξ_t is that they often exhibit *finite* convergence. Indeed, there is an (easy to check) criterion, known as the *flatness condition*, which permits to conclude that the level t bound is exact, i.e., $\xi_t = \text{val}$, and to extract a finite atomic optimal solution of GMP. This flatness condition, see (20) below, goes back to work of Curto and Fialkow [19, 20]. We also refer, e.g., to [46, 49] for a detailed

exposition of the following result. For details on how to extract an atomic optimal solution under the flatness condition (20), we refer to [37, 49].

Theorem 4 [19, 20] *Consider the set K from (2) and set $d_K = \max\{1, \lceil \deg(g_j)/2 \rceil :$ $j \in [m]\}$. Let $t \in \mathbb{N}$ such that $2t \geq \max\{\deg(f_i) : 0 \leq i \leq N\}$ and $t \geq d_K$. Assume $L \in \mathbb{R}[x]_{2t}^*$ is an optimal solution to the program (6) defining the parameter ξ_t and it satisfies the following* flatness condition:

$$\operatorname{rank} L([x]_s [x]_s^T) = \operatorname{rank} L([x]_{s-d_K} [x]_{s-d_K}^T) =: r$$
$$\text{for some integer } s \text{ such that } d_K \leq s \leq t. \qquad (20)$$

Then, equality $\xi_t = $ val holds, and problem (1) has an optimal solution μ that is finite atomic and supported on r points in K.

The above result naturally applies also to the sparse reformulation (11) of GMP and to the sparse hierarchy ξ_t^{isp} in (12). Indeed, it suffices to apply Theorem 4 to each of the linear functionals L_k and to check whether L_k satisfies the corresponding flatness criterion. We adapt the result to this setting for concreteness.

Corollary 5 *Consider the sets K in (2) and K_k in (10) and define the parameter $d_{K_k} = \max\{1, \lceil \deg((g_j)_{|V_k})/2 \rceil : j \in [m]\}$ for $k \in [p]$. Let $t \in \mathbb{N}$ such that $2t \geq \max\{\deg(f_i) : 0 \leq i \leq m\}$ and $t \geq \max\{d_{K_k} : k \in [p]\}$. Assume $(L_1, \ldots, L_p)$ is an optimal solution to the program (12) defining ξ_t^{isp} and it satisfies the flatness condition: for each $k \in [p]$ there exists an integer s_k such that $d_{K_k} \leq s_k \leq t$ and the following holds*

$$\operatorname{rank} L_k([x(V_k)]_{s_k} [x(V_k)]_{s_k}^T) = \operatorname{rank} L_k([x(V_k)]_{s_k-d_{K_k}} [x(V_k)]_{s_k-d_{K_k}}^T) =: r_k. \quad (21)$$

Then, equality $\xi_t^{\mathrm{isp}} = \mathrm{val}^{\mathrm{isp}}(= \mathrm{val})$ holds, and problem (11) has an optimal solution $(\mu_1, \ldots, \mu_p)$, where each μ_k is finite atomic and supported on r_k atoms in K_k for each $k \in [p]$.

Note that, for the application to the completely positive rank and the nonnegative rank, all involved polynomials in the corresponding instances of GMP are quadratic, so that $d_K = d_{K_k} = 1$ and the smallest relaxation level that can be considered is $t = 1$. For the application to the cp-rank, if the flatness condition holds for an optimal solution for the parameter $\xi_t^{\mathrm{cp}}(A)$ (or for the parameter $\xi_t^{\mathrm{cp,isp}}(A)$), then the parameter is equal to $\tau_{\mathrm{cp}}(A)$ and one can extract a cp-factorization of A. In this way one finds an explicit factorization of A and thus an upper bound on its cp-rank. In this case, if the computed value of $\tau_{\mathrm{cp}}(A)$ is equal to the number of recovered atoms, this certifies that $\tau_{\mathrm{cp}}(A)$ is equal to the cp-rank and the recovered cp-decomposition of A is an optimal one. We will illustrate this on some examples in Sect. 4.3.2. In the same way, for the application to the nonnegative rank, if the flatness condition holds for an optimal solution for the parameter $\xi_t^{+}(M)$ (or for the parameter $\xi_t^{+,\mathrm{isp}}(M)$), then the parameter is equal to $\tau_{+}(M)$ and one can extract a nonnegative factorization of M.

3 Ideal-sparsity for GMP

In this section we investigate how ideal-sparsity can be exploited for the GMP (1). First, we consider in Sect. 3.1 the ideal-sparse reformulation (11) and the corresponding ideal-sparse bounds, and, after that, we mention in Sect. 3.2 how this relates to the more classic approach based on exploiting correlative sparsity.

3.1 Ideal-sparse moment relaxations

Consider the GMP (1), where the set K is defined as in (2). As in Sect. 1, we consider the graph $G = (V, E)$, whose maximal cliques are denoted $V_1, \ldots, V_p$, and we define the sets $\widehat{K_k} \subseteq K \subseteq \mathbb{R}^n$ (as in (8)) and their projections $K_k \subseteq \mathbb{R}^{|V_k|}$ (as in (10)). Recall from (9) that $K = \widehat{K_1} \cup \ldots \cup \widehat{K_p}$. Then, one can define the (sparse) version (11) of GMP. As observed above, while problem (1) has a single measure variable μ whose support is contained in $K \subseteq \mathbb{R}^n$, problem (11) involves p measure variables $\mu_1, \ldots, \mu_p$, where μ_k is supported on the set $K_k \subseteq \mathbb{R}^{|V_k|}$, thus a smaller dimensional space. We now show that both formulations (1) and (11) are equivalent.

Proposition 6 *Problems* (1) *and* (11) *are equivalent, i.e., their optimum values are equal:* $\mathrm{val} = \mathrm{val}^{\mathrm{isp}}$.

Proof First, we show $\mathrm{val} \leq \mathrm{val}^{\mathrm{isp}}$. For this, assume $(\mu_1, \ldots, \mu_p)$ is feasible for problem (11). Consider the measure μ on $\mathbb{R}^{|V|}$, defined by $\int f d\mu = \sum_{k=1}^{p} \int_{K_k} f_{|V_k} d\mu_k$ for any measurable function f on $\mathbb{R}^{|V|}$. We have $\mathrm{Supp}(\mu) \subseteq K$. Indeed, $\int_K f d\mu = \int f \chi^K d\mu = \sum_k \int_{K_k} f_{|V_k} \chi_{|V_k}^K d\mu_k = \sum_k \int_{K_k} f_{|V_k} d\mu_k = \int f d\mu$, since $\chi_{|V_k}^K(y) = \chi^K(y, 0_{V \setminus V_k}) = 1$ for all $y \in K_k$ as $(y, 0_{V \setminus V_k}) \in \widehat{K_k} \subseteq K$. Then, μ is feasible for (1), with the same objective value as $(\mu_1, \ldots, \mu_p)$, which shows $\mathrm{val} \leq \mathrm{val}^{\mathrm{isp}}$.

We now show the reverse inequality $\mathrm{val}^{\mathrm{isp}} \leq \mathrm{val}$. For this, assume μ is feasible for (1). We now define a feasible solution $(\mu_1, \ldots, \mu_p)$ to (11), with the same objective value as μ. For $k \in [p]$, define the set

$$\Lambda_k = \{x \in K : \mathrm{Supp}(x) \subseteq V_k, \ \mathrm{Supp}(x) \nsubseteq V_h \text{ for } 1 \leq h \leq k - 1\}.$$

As each $x \in K$ has its support contained in some V_k, it follows that the sets $\Lambda_1, \ldots, \Lambda_p$ form a disjoint partition of K. Note that $\Lambda_k \subseteq \widehat{K_k}$ and thus $x(V_k) \in K_k$ for any $x \in \Lambda_k$. Consider the measure μ_k on $\mathbb{R}^{|V_k|}$, defined by $\int f d\mu_k = \int_{\Lambda_k} f(x(V_k)) d\mu(x)$ for any measurable function f on $\mathbb{R}^{|V_k|}$. Then, $\mathrm{Supp}(\mu_k) \subseteq K_k$, since $\int_{K_k} f d\mu_k = \int f \chi^{K_k} d\mu_k = \int_{\Lambda_k} f(x(V_k)) \chi^{K_k}(x(V_k)) d\mu(x) = \int_{\Lambda_k} f(x(V_k)) d\mu(x) = \int f d\mu_k$, as $\chi^{K_k}(x(V_k)) = 1$ for all $x \in \Lambda_k$. Next, we show that $\int p d\mu = \sum_k \int p_{|V_k} d\mu_k$ for any measurable function $p : \mathbb{R}^{|V|} \to \mathbb{R}$. Indeed, as the sets $\Lambda_1, \ldots, \Lambda_p$ disjointly partition the set K, we have $\int p d\mu = \int_K p d\mu = \sum_k \int_{\Lambda_k} p d\mu$. Combining with $\int_{\Lambda_k} p(x) d\mu(x) = \int_{\Lambda_k} p_{|V_k}(x(V_k)) d\mu(x) = \int_{K_k} p_{|V_k} d\mu_k$, gives the desired identity $\int p d\mu = \sum_k \int p_{|V_k} d\mu_k$. Therefore, $(\mu_1, \ldots, \mu_p)$ is a feasible solution to (11) with the same value as μ, which shows $\mathrm{val}^{\mathrm{isp}} \leq \mathrm{val}$. □

Based on the reformulation (11), we can define the *ideal-sparse* moment relaxation (12) for problem (1), which we repeat here for convenience: for any integer $t \geq 1$,

$$\xi_t^{\text{isp}} := \inf \left\{ \sum_{k=1}^{p} L_k(f_{0|V_k}) : L_k \in \mathbb{R}[x(V_k)]_{2t}^* \ (k \in [p]), \right.$$
$$\sum_{k=1}^{p} L_k(f_{i|V_k}) = a_i \ (i \in [N]), \quad (22)$$
$$\left. L_k \geq 0 \text{ on } \mathcal{M}(\mathbf{g}_{|V_k})_{2t} \ (k \in [p]) \right\}.$$

This hierarchy provides bounds for val that are at least at good as the bounds ξ_t from (6).

Theorem 7 *For any integer $t \geq 1$ we have $\xi_t \leq \xi_t^{\text{isp}} \leq$ val. In addition, if $\mathcal{M}(\mathbf{g})$ is Archimedian and (7) holds, then $\lim_{t \to \infty} \xi_t^{\text{isp}} =$ val.*

Proof Clearly, $\xi_t^{\text{isp}} \leq$ val$^{\text{isp}}$, which, combined with Proposition 6, gives $\xi_t^{\text{isp}} \leq$ val. We now show $\xi_t \leq \xi_t^{\text{isp}}$. For this, assume $(L_1, \ldots, L_p)$ is feasible for (22). Define $L \in \mathbb{R}[x]_{2t}^*$ by setting $L(p) = \sum_{k=1}^{p} L_k(p_{|V_k})$ for any $p \in \mathbb{R}[x]_{2t}$. By construction, $L(f_i) = \sum_k L_k(f_{i|V_k})$ for $0 \leq i \leq m$, so that $L(f_i) = a_i$ for $i \in [m]$), and $L \geq 0$ on $\mathcal{M}(\mathbf{g})$. For each $\{i, j\} \in \overline{E}$ and $k \in [p]$, we have $\{i, j\} \not\subseteq V_k$ and thus $(x_i x_j)_{|V_k}$ is identically zero; hence, for any $u \in \mathbb{R}[x]_{2t-2}$, we have $L(u x_i x_j) = \sum_k L_k(u_{|V_k}(x_i x_j)_{|V_k}) = 0$. Hence, L is feasible for (6) with the same objective value as $(L_1, \ldots, L_p)$, which shows $\xi_t \leq \xi_t^{\text{isp}}$. Convergence of ξ_t^{isp} to val follows from the just proven fact that $\xi_t \leq \xi_t^{\text{isp}}$ and from Theorem 1, which implies $\lim_{t \to \infty} \xi_t =$ val under the stated assumptions. □

Observe that in Theorem 7 no structural chordality property needs to be assumed on the cliques $V_1, \ldots, V_p$ of the graph G. In other words, the cliques $V_1, \ldots, V_p$ need not satisfy the running intersection property (see (24) below), which is a characterizing property of chordal graphs that is often used in sparsity exploiting techniques like correlative sparsity. In Sect. 3.2 below, we will comment about the link between the ideal-sparsity approach presented here and the more classical correlative sparsity approach that can be followed when considering a chordal extension $\widehat{G}$ of the graph G.

As mentioned earlier in the introduction, the sparse bounds ξ_t^{isp} present a double advantage compared to the dense bounds ξ_t: they are at least as good (and often strictly better), and their computation is potentially faster since the sets V_k can be much smaller than the full set V. We will see later examples illustrating this. On the other hand, a possible drawback is that the number of maximal cliques of G could be large. Indeed, it is well-known that the number of maximal cliques can be exponential in the number of nodes (this is the case, e.g., when G is a complete graph on $2n$ nodes with a deleted perfect matching). A possible remedy is to consider a graph $\widetilde{G} = (V, \widetilde{E})$ containing G as a subgraph, i.e., such that $E \subseteq \widetilde{E}$. Then, let $\widetilde{V}_1, \ldots, \widetilde{V}_{\widetilde{p}}$ denote the maximal cliques of $\widetilde{G}$, whose number $\widetilde{p}$ satisfies $\widetilde{p} \leq p$, since each maximal clique of G is contained in a maximal clique of $\widetilde{G}$. One can define the corresponding ideal-sparse moment

hierarchy of bounds, denoted $\widetilde{\xi}_t^{\mathrm{isp}}$, which involves $\widetilde{p}$ measure variables supported on the sets $\widetilde{V}_1, \ldots, \widetilde{V}_{\widetilde{p}}$ (instead of the sets $V_1, \ldots, V_p$). However, as $\widetilde{V}_h$ may contain some non-edge of G, one now needs to still impose an ideal condition on each linear functional $\widetilde{L}_h$ acting on $\mathbb{R}[x(\widetilde{V}_h)]$ ($h \in [\widetilde{p}]$). Namely, the parameter $\widetilde{\xi}_t^{\mathrm{isp}}$ is defined as

$$
\widetilde{\xi}_t^{\mathrm{isp}} := \inf \left\{ \sum_{h=1}^{\widetilde{p}} \widetilde{L}_h(f_{0|\widetilde{V}_h}) : \widetilde{L}_h \in \mathbb{R}[x(\widetilde{V}_h)]_{2t}^* \ (h \in [\widetilde{p}]), \right.
$$

$$
\sum_{h=1}^{\widetilde{p}} \widetilde{L}_h(f_{i|\widetilde{V}_h}) = a_i \ (i \in [N]),
$$

$$
\widetilde{L}_h \geq 0 \text{ on } \mathcal{M}(\mathbf{g}_{|\widetilde{V}_h})_{2t} \ (h \in [\widetilde{p}]),
$$

$$
\widetilde{L}_h(x_i x_j x^\alpha) = 0 \ (\alpha \in \mathbb{N}_{2t-2}^n,
$$

$$
\left. \mathrm{Supp}(\alpha) \subseteq \widetilde{V}_h, \ \{i, j\} \subseteq \widetilde{V}_h, \ \{i, j\} \in \overline{E}) \right\}.
$$

(23)

Note that this parameter interpolates between the dense and sparse parameters: indeed, $\widetilde{\xi}_t^{\mathrm{isp}} = \xi_t^{\mathrm{isp}}$ if $\widetilde{G} = G$, and $\widetilde{\xi}_t^{\mathrm{isp}} = \xi_t$ if $\widetilde{G} = K_n$ is the complete graph. Accordingly, we have the following inequalities among the parameters.

Lemma 8 *Assume $\widetilde{G}$ contains G as a subgraph. For any integer $t \geq 1$ we have* $\xi_t \leq \widetilde{\xi}_t^{\mathrm{isp}} \leq \xi_t^{\mathrm{isp}}$.

Proof The proof for the inequality $\xi_t \leq \widetilde{\xi}_t^{\mathrm{isp}}$ is analogous to the proof of $\xi_t \leq \xi_t^{\mathrm{isp}}$ in Theorem 7. We now show $\widetilde{\xi}_t^{\mathrm{isp}} \leq \xi_t^{\mathrm{isp}}$. For this, assume $(L_1, \ldots, L_p)$ is feasible for the parameter ξ_t^{isp}. As each clique V_k of G is contained in some clique $\widetilde{V}_h$ of $\widetilde{G}$, there exists a partition $[p] = A_1 \cup \ldots \cup A_{\widetilde{p}}$ such that $V_k \subseteq \widetilde{V}_h$ for all $k \in A_h$ and $h \in [\widetilde{p}]$. For $h \in [\widetilde{p}]$, we define $\widetilde{L}_h \in \mathbb{R}[x(\widetilde{V}_h)]_{2t}^*$ by setting $\widetilde{L}_h(p) = \sum_{k \in A_h} L_k(p_{|V_k})$ for $p \in \mathbb{R}[x(\widetilde{V}_h)]_{2t}$. Then, one can easily verify that $(\widetilde{L}_1, \ldots, \widetilde{L}_{\widetilde{p}})$ provides a feasible solution for $\widetilde{\xi}_t^{\mathrm{isp}}$, with the same objective value as $(L_1, \ldots, L_p)$. Let us only check the ideal constraint. For this assume $\{i, j\} \cup \mathrm{Supp}(\alpha) \subseteq \widetilde{V}_h$ and $\{i, j\} \in \overline{E}$. Then, $\{i, j\}$ is not contained in any clique V_k of G and thus $L_k((x_i x_j x^\alpha)_{|V_k}) = 0$ for all $k \in [p]$, which directly implies $\widetilde{L}_h(x_i x_j x^\alpha) = 0$. □

Remark 9 There may be a trade-off to be made between the parameter ξ_t^{isp}, which fully exploits the sparsity of G (and provides a possibly better bound), and the parameter $\widetilde{\xi}_t^{\mathrm{isp}}$, which only partially exploits the sparsity, depending on the choice of the extension $\widetilde{G}$ of G. Namely, the parameter ξ_t^{isp} may involve many cliques of smaller sizes, while the parameter $\widetilde{\xi}_t^{\mathrm{isp}}$ involves less cliques but with larger sizes. If one cares to have a small number of cliques, then one can (but is not required to) consider for $\widetilde{G}$ a chordal extension $\widehat{G}$ of G, in which case the number of maximal cliques is at most the number of nodes.

In our numerical experiments for matrix factorization ranks we will consider only the two extreme cases of the dense and ideal-sparse parameters ξ_t and ξ_t^{isp}. For most of the matrices considered the number of maximal cliques seems indeed not to play a significant role. However, when this number becomes too large, one may have to consider alternative intermediate parameters (see Sect. 6 for a brief discussion).

3.2 Bounds based on correlative sparsity

In this section we compare the ideal-sparse approach with the more classic one based on exploiting correlative sparsity. The setting of correlative sparsity is usually applied to a polynomial optimization problem, where each of the polynomials arising as a constraint involves only a subset of the variables (indexed, say, by one of the subsets $\widehat{V}_1, \ldots, \widehat{V}_{\widehat{p}}$) and the objective polynomial is a sum of such polynomials. Then, one can define more economical relaxations that respect this sparsity pattern. In the case when the sets $\widehat{V}_1, \ldots, \widehat{V}_{\widehat{p}}$ respect the so-called RIP property (see (24) below) (and under some Archimedean condition), these hierarchies enjoy asymptotic convergence properties analogous to the dense hierarchies; see [35, 44] for details and also [51] for general background on correlative sparsity. We now explain how correlative sparsity applies to the instance of GMP considered in this paper.

As before, we assume K is contained in the variety of the ideal I_E, generated by the monomials $x_i x_j$ corresponding to the nonedges of the graph $G = (V, E)$. In the ideal-sparsity approach we considered a measure variable for each maximal clique of G. However, the number of maximal cliques of G can be large, which could represent a drawback for this approach.

An alternative is to consider a *chordal extension* $\widehat{G} = (V, \widehat{E})$ of G, that is, a chordal graph $\widehat{G}$ containing G as a subgraph, i.e., such that $E \subseteq \widehat{E}$. Then, as a well-known property of chordal graphs, $\widehat{G}$ has at most n distinct maximal cliques. Let $\widehat{V}_1, \ldots, \widehat{V}_{\widehat{p}}$ denote the maximal cliques of $\widehat{G}$, so $\widehat{p} \leq n$. As one of the many equivalent definitions of chordal graphs, it is known that the maximal cliques $\widehat{V}_1, \ldots, \widehat{V}_{\widehat{p}}$ satisfy (possibly after reordering) the so-called *running intersection property* (RIP):

$$\forall k \in \{2, \ldots, \widehat{p}\} \; \exists j \in \{1, \ldots, k-1\} \; \text{ such that } \; \widehat{V}_k \cap (\widehat{V}_1 \cup \ldots \cup \widehat{V}_{k-1}) \subseteq \widehat{V}_j. \quad (24)$$

See, e.g., [25] for details. As we explain below, it turns out that one can 'transport' the chordal sparsity structure of the graph $\widehat{G}$ to the moment matrices involved in the definition of the dense bound ξ_t in (6).

To see this, let us first rewrite the parameter ξ_t more concretely as a semidefinite program. For convenience, set $d_j := \lceil \deg(g_j)/2 \rceil$ for $j \in [m]$. Then, following the discussion in Sect. 2.1, the parameter ξ_t can be expressed as

$$\begin{aligned}
\xi_t = \inf\{ & L(f_0) : L \in \mathbb{R}[x]_{2t}^*, \; L(f_i) = a_i \; (i \in [N]), \\
& L([x]_t[x]_t^T) \succeq 0, \; L(g_j[x]_{t-d_j}[x]_{t-d_j}^T) \succeq 0 \; (j \in [m]), \\
& L = 0 \text{ on } I_{E,2t}, \text{ i.e., } L(x_i x_j x^\alpha) = 0 \; (\{i, j\} \in \overline{E}, \; \alpha \in \mathbb{N}_{2t-2}^n)\}.
\end{aligned} \quad (25)$$

For fixed $t \in \mathbb{N}$, define the sets

$$\mathcal{I}_{k,t} = \{\alpha \in \mathbb{N}_t^n : \mathrm{Supp}(\alpha) \subseteq \widehat{V}_k\} \subseteq \mathbb{N}_t^n \; (k \in [\widehat{p}]), \quad \mathcal{I}_t = \bigcup_{k=1}^{\widehat{p}} \mathcal{I}_{k,t} \subseteq \mathbb{N}_t^n. \quad (26)$$

Lemma 10 *Assume $L \in \mathbb{R}[x]_{2t}^*$ satisfies $L = 0$ on $I_{E,2t}$. Then, $L(x^\alpha x^\beta) = 0$ for any $\alpha, \beta \in \mathbb{N}_t^n$ such that $\{\alpha, \beta\}$ is not contained in any of the sets $\mathcal{I}_{1,t}, \ldots, \mathcal{I}_{\widehat{p},t}$.*

Proof Assume there is no index $k \in [\widehat{p}]$ such that $\{\alpha, \beta\} \subseteq \mathcal{I}_{k,t}$. Then, Supp$(\alpha + \beta)$ is not a clique in G, for otherwise it would be contained in some $\widehat{V}_k$, implying Supp(α), Supp$(\beta) \subseteq \widehat{V}_k$ and thus $\alpha, \beta \in \mathcal{I}_{k,t}$, yielding a contradiction. As Supp$(\alpha + \beta)$ is not a clique in G, it contains a pair $\{i, j\} \in \overline{E}$, which implies $x^\alpha x^\beta \in I_{E,2t}$ and thus $L(x^\alpha x^\beta) = 0$. $\square$

In view of Lemma 10, in the definition of ξ_t in (25), one may restrict the matrix $L([x]_t[x]_t^T)$ to its principal submatrix indexed by $\mathcal{I}_t$, since any row/column indexed by $\alpha \in \mathbb{N}_t^n \setminus \mathcal{I}_t$ is identically zero. Moreover, $L(x^\alpha x^\beta) \neq 0$ implies $\{\alpha, \beta\} \subseteq \mathcal{I}_{k,t}$ for some $k \in [\widehat{p}]$. In other words, the support graph of the matrix $L([x]_t[x]_t^T)$ is contained in the graph with vertex set $\mathcal{I}_t$, whose maximal cliques are the sets $\mathcal{I}_{1,t}, \ldots, \mathcal{I}_{\widehat{p},t}$. The next lemma shows that the RIP property also holds for the sets $\mathcal{I}_{1,t}, \ldots, \mathcal{I}_{\widehat{p},t}$. Therefore, the moment matrix $M_t(L) = L([x]_t[x]_t^T)$ has a correlative sparsity pattern, which it inherits from the chordal extension $\widehat{G}$ of G.

Lemma 11 *The sets $\mathcal{I}_{1,t}, \ldots, \mathcal{I}_{\widehat{p},t}$ satisfy the RIP property:*

$$\forall q \in \{2, \ldots, \widehat{p}\} \ \exists k \in \{1, \ldots, q-1\} \quad such\ that \quad \mathcal{I}_{q,t} \cap (\mathcal{I}_{1,t} \cup \ldots \cup \mathcal{I}_{q-1,t}) \subseteq \mathcal{I}_{k,t}.$$
(27)

Proof Let $q \in \{2, \ldots, \widehat{p}\}$ and assume by way of contradiction that there exists no $k \in [q-1]$ for which $\mathcal{I}_{q,t} \cap (\mathcal{I}_{1,t} \cup \ldots \cup \mathcal{I}_{q-1,t}) \subseteq \mathcal{I}_k$ holds. Then, for each $k \in [q-1]$, there exists $\alpha^k \in \mathcal{I}_{q,t} \cap (\mathcal{I}_{1,t} \cup \ldots \cup \mathcal{I}_{q-1,t}) \setminus \mathcal{I}_{k,t}$ and thus there exists $i_k \in V \setminus \widehat{V}_k$ such that $\alpha_{i_k}^k \geq 1$. As $\alpha^k \in \mathcal{I}_{q,t}$ and $\alpha_{i_k}^k \geq 1$, it follows that $i_k \in \widehat{V}_q$. In addition, $\alpha^k \in \mathcal{I}_{j,t}$ for some $j \in [q-1]$. Again, as $\alpha_{i_k}^k \geq 1$, it follows that $i_k \in \widehat{V}_j$. This shows that

$$i_k \in \widehat{V}_q \cap (\widehat{V}_1 \cup \ldots \cup \widehat{V}_{q-1}) \quad \text{for all } k \in [q-1].$$

By the RIP property (24) for $\widehat{V}_1, \ldots, \widehat{V}_p$, there exists $q_0 \in [q-1]$ such that $\widehat{V}_q \cap (\widehat{V}_1 \cup \ldots \cup \widehat{V}_{q-1}) \subseteq \widehat{V}_{q_0}$. Therefore, $i_k \in \widehat{V}_{q_0}$ for all $k \in [q-1]$. As $i_k \notin \widehat{V}_k$, this implies that $q_0 \neq k$ for all $k \in [q-1]$, and we reach a contradiction. $\square$

The above extends to the localizing matrices $L(g_j[x]_{t-d_j}[x]_{t-d_j}^T)$ for $j \in [m]$. In the same way, one may restrict the matrix $L(g_j[x]_{t-d_j}[x]_{t-d_j}^T)$ to its principal submatrix indexed by $\mathcal{I}_{t-d_j}$ and its support graph is contained in the graph with vertex set $\mathcal{I}_{t-d_j}$, whose maximal cliques are the sets $\mathcal{I}_{1,t-d_j}, \ldots, \mathcal{I}_{\widehat{p},t-d_j}$. Moreover, there is a correlative sparsity pattern on the matrix $L(g_j[x]_{t-d_j}[x]_{t-d_j}^T)$ $(0 \leq j \leq m)$, which is inherited from the chordal structure of $\widehat{G}$.

Therefore, one may apply Theorem 12 below to get a more economical reformulation of ξ_t. Indeed, by Theorem 12, one may write $L(g_j[x]_{t-d_j}[x]_{t-d_j}^T) = \sum_{k=1}^{\widehat{p}} Z_{j,k}$, where $Z_{j,k}$ is obtained from a matrix indexed by the set $\mathcal{I}_{k,t-d_j}$ by padding it with zero entries, and replace the condition $L(g_j[x]_{t-d_j}[x]_{t-d_j}^T) \succeq 0$ by the conditions $Z_{j,1}, \ldots, Z_{j,\widehat{p}} \succeq 0$. The advantage is that requiring $Z_{j,k} \succeq 0$ boils down to checking positive semidefiniteness of a potentially much smaller matrix, indexed by $\mathcal{I}_{k,t-d_j}$.

Hence, this allows to replace one (large) positive semidefinite matrix by several smaller positive semidefinite matrices. While this method offers a more economical way for computing the dense parameter ξ_t, it is nevertheless inferior to the ideal-sparse approach described in the previous section. Recall in particular Remark 9, where we indicated how to construct a sparse parameter $\widetilde{\xi}_t^{\mathrm{isp}}$, which could also be based on using a chordal extension $\widehat{G}$ of G, but superior in quality as $\xi_t \leq \widetilde{\xi}_t^{\mathrm{isp}}$.

Theorem 12 ([2]) *Consider a positive semidefinite matrix $X \in \mathcal{S}_+^n$ whose support graph is contained in a chordal graph $\widehat{G}$, with maximal cliques $\widehat{V}_1, \ldots, \widehat{V}_{\widehat{p}}$. Then, there exist positive semidefinite matrices $Y_k \in \mathcal{S}_+^{\widehat{V}_k}$ ($k \in [\widehat{p}]$) such that $X = \sum_{k=1}^{\widehat{p}} Z_k$, where $Z_k = Y_k \oplus 0_{V \setminus \widehat{V}_k, V \setminus \widehat{V}_k} \in \mathcal{S}_+^n$ is obtained by padding Y_k with zeros.*

As a final observation, another possibility to exploit the above correlative sparsity structure would be simply to replace in the definition of ξ_t in program (6) each condition $L(g_j[x]_{t-d_j}[x]_{t-d_j}) \succeq 0$ by $\widehat{p}$ smaller matrix conditions $L(g_{j|\widehat{V}_k}[x(\widehat{V}_k)]_{t-d_j}[x(\widehat{V}_k)]_{t-d_j}) \succeq 0$ for $k \in [\widehat{p}]$. In other words, if $L_{|V_k}$ denotes the restriction of L to the polynomials in variables indexed by $\widehat{V}_k$, then we replace the condition $L \geq 0$ on $\mathcal{M}(\mathbf{g})_{2t}$ by the conditions $L_{|\widehat{V}_k} \geq 0$ on $\mathcal{M}(\mathbf{g}_{|\widehat{V}_k})_{2t}$ for each $k \in [\widehat{p}]$. In this way we obtain another parameter, denoted by ξ_t^{csp}, that is weaker than ξ_t and thus satisfies

$$\xi_t^{\mathrm{csp}} \leq \xi_t \leq \widetilde{\xi}_t^{\mathrm{isp}} \leq \xi_t^{\mathrm{isp}}.$$

Recall $\widetilde{\xi}_t^{\mathrm{isp}}$ is the parameter from (23) obtained when selecting an extension $\widetilde{G}$ of G, including, for instance, selecting a chordal extension $\widetilde{G} = \widehat{G}$.

4 Application to the completely positive rank

In this section we investigate how ideal-sparsity can be exploited to design bounds on the completely positive rank. We define the corresponding hierarchies of lower bounds on the cp-rank and indicate their relations to other known bounds in the literature.

4.1 Ideal-sparse lower bounds on the cp-rank

Consider a symmetric nonnegative matrix $A \in \mathcal{S}^n$ and assume $A_{ii} \neq 0$ for all $i \in V$ (to avoid trivialities). Then, its cp-rank, denoted $\mathrm{rank}_{\mathrm{cp}}(A)$, is the smallest integer $r \in \mathbb{N}$ for which A admits a decomposition of the form $A = \sum_{\ell=1}^r a_\ell a_\ell^T$ with $a_\ell \geq 0$ (setting $r = \infty$ if no such decomposition exists, when A is not completely positive). Fawzi and Parrilo [29] introduced the parameter $\tau_{\mathrm{cp}}(A)$ from (13), as a convexification of the cp-rank, whose definition is repeated for convenience:

$$\tau_{\mathrm{cp}}(A) := \min \left\{ \lambda : \frac{1}{\lambda} A \in \mathrm{conv}\{xx^T : x \in \mathbb{R}_+^n, \ A - xx^T \succeq 0, \ A \geq xx^T\} \right\}.$$

Clearly, we have $\tau_{\mathrm{cp}}(A) \leq \mathrm{rank}_{\mathrm{cp}}(A)$. As was already indicated in Sect. 1, the parameter $\tau_{\mathrm{cp}}(A)$ can be reformulated as an instance of problem (1) with an ideal-sparsity

structure inherited from the matrix A. For this, recall $G_A = (V = [n], E_A)$ denotes the support graph of A, where E_A consists of all pairs $\{i, j\}$ with $i \neq j \in V$ and $A_{ij} \neq 0$ (as in (14)), and recall the definition of the semialgebraic set K_A from (15). As shown in Lemma 2, $\tau_{cp}(A)$ can be reformulated as an instance of GMP:

$$\tau_{cp}(A) = \inf_{\mu \in \mathcal{M}(\mathbb{R}^n)} \left\{ \int_{K_A} 1 d\mu : \int_{K_A} x_i x_j d\mu = A_{ij} \ (i, j \in V), \ \mathrm{Supp}(\mu) \subseteq K_A \right\}.$$

Dense hierarchies for cp-rank.   Based on the above reformulation of $\tau_{cp}(A)$, for any integer $t \geq 1$, let us define the following parameter (as special instance of (6)):

$$\xi_t^{cp}(A) = \min \left\{ L(1) : L \in \mathbb{R}[x]_{2t}^*, \right. \tag{28}$$

$$L(x_i x_j) = A_{ij} \ (i, j \in V), \tag{29}$$

$$L([x]_t [x]_t^T) \succeq 0, \tag{30}$$

$$L((\sqrt{A_{ii}} x_i - x_i^2)[x]_{t-1}[x]_{t-1}^T) \succeq 0 \ \text{for } i \in V, \tag{31}$$

$$L((A_{ij} - x_i x_j)[x]_{t-1}[x]_{t-1}^T) \succeq 0 \ \text{for } \{i, j\} \in E_A, \tag{32}$$

$$L(x_i x_j [x]_{2t-2}) = 0 \ \text{for } \{i, j\} \in \overline{E}_A, \tag{33}$$

$$L((A - xx^T) \otimes [x]_{t-1}[x]_{t-1}^T) \succeq 0. \tag{34}$$

We first indicate how this parameter relates to other similar moment-based bounds considered in the literature, in particular in [33] and [34]. Note that, due to the presence of the (ideal) constraints (33), the constraint (32) trivially holds for any pair $\{i, j\} \in \overline{E}_A$. If we omit the ideal constraint (33) and impose the constraint (32) for *all* pairs $\{i, j\}$ with $i \neq j \in V$, then we obtain a parameter investigated in [34], denoted here as $\xi_{t,(2022)}^{cp}(A)$. The parameter $\xi_{t,(2022)}^{cp}(A)$ strengthens an earlier parameter $\xi_{t,(2019)}^{cp}(A)$ introduced in [33], whose definition follows by replacing in the definition of $\xi_{t,(2022)}^{cp}(A)$ the constraint (34) by the weaker constraint

$$L((xx^T)^{\otimes \ell}) \preceq A^{\otimes \ell} \ \text{for } \ell \in [t]. \tag{35}$$

So, for any $t \geq 1$, we have

$$\xi_{t,(2019)}^{cp}(A) \leq \xi_{t,(2022)}^{cp}(A) \leq \xi_t^{cp}(A).$$

Since the bounds $\xi_{t,(2019)}^{cp}(A)$ were shown to converge asymptotically to $\tau_{cp}(A)$ in [33], the same holds for the bounds $\xi_t^{cp}(A)$. Note that the convergence of the latter bounds also follows directly from Theorem 1.

As mentioned in [33], there are more constraints that can be added to the above program and still lead to a lower bound on the cp-rank (in fact on $\tau_{cp}(A)$). In particular, exploiting the fact that the variables x_i should be nonnegative, one may add the

constraints

$$L([x]_{2t}) \geq 0, \tag{36}$$

$$L((\sqrt{A_{ii}}x_i - x_i^2)[x]_{2t-2}) \geq 0 \text{ for } i \in V, \tag{37}$$

$$L(A_{ij} - x_i x_j)[x]_{2t-2}) \geq 0 \text{ for } \{i, j\} \in E_A. \tag{38}$$

One may also add other localizing constraints, such as

$$L(x_i x_j [x]_{t-1}[x]_{t-1}^T) \succeq 0 \text{ for } \{i, j\} \in E_A. \tag{39}$$

Note that the constraints (39) are redundant at the smallest level $t = 1$. Note also that one could add a similar constraint replacing $x_i x_j$ by any monomial. We use the notation $\xi_{t,\dagger}^{cp}(A)$ to denote the parameter obtained by adding (38) to the program defining $\xi_t^{cp}(A)$. Define analogously $\xi_{t,(2019),\dagger}^{cp}(A)$ by adding (38) to $\xi_{t,(2019)}^{cp}(A)$, so that we have

$$\xi_{t,(2019),\dagger}^{cp}(A) \leq \xi_{t,\dagger}^{cp}(A).$$

As we will see in relation (52) below, the bound $\xi_{2,(2019),\dagger}^{cp}(A)$ is at least as good as $\operatorname{rank}(A)$, an obvious lower bound on $\operatorname{rank}_{cp}(A)$. Let $\xi_{t,\ddagger}^{cp}(A)$ denote the strengthening of $\xi_{t,\dagger}^{cp}(A)$ by adding constraints (36), (37), and (39), so that we have $\xi_t^{cp}(A) \leq \xi_{t,\dagger}^{cp}(A) \leq \xi_{t,\ddagger}^{cp}(A)$.

Ideal-sparse hierarchies for cp-rank. We now consider the ideal-sparse bounds for the cp-rank, which further exploit the ideal-sparsity pattern of A. For this, let $V_1, \ldots, V_p$ denote the maximal cliques of the graph G_A and, for $t \geq 1$, define the following parameter (as special instance of (12)):

$$\xi_t^{cp,isp}(A) = \min \left\{ \sum_{k=1}^p L_k(1) : L_k \in \mathbb{R}[x(V_k)]_{2t}^* \ (k \in [p]), \right. \tag{40}$$

$$\sum_{k \in [p]:i,j \in V_k} L_k(x_i x_j) = A_{ij} \ (i, j \in V), \tag{41}$$

$$L_k([x(V_k)]_t[x(V_k)]_t^T) \succeq 0 \ (k \in [p]), \tag{42}$$

$$L_k((\sqrt{A_{ii}}x_i - x_i^2)[x(V_k)]_{t-1}[x(V_k)]_{t-1}^T) \succeq 0 \ \text{for } i \in V_k, \ k \in [p], \tag{43}$$

$$L_k((A_{ij} - x_i x_j)[x(V_k)]_{t-1}[x(V_k)]_{t-1}^T) \succeq 0 \ \text{for } i \neq j \in V_k, \ k \in [p], \tag{44}$$

$$L_k((A - xx^T) \otimes [x(V_k)]_{t-1}[x(V_k)]_{t-1}^T) \succeq 0, \ \text{ for } k \in [p]. \tag{45}$$

Here, in equation (45), it is understood that, for a given $k \in [p]$, in the matrix $A - xx^T$ one sets the entries of x indexed by $V \setminus V_k$ to zero. As a direct application of Theorem

7, we have

$$\xi_t^{\mathrm{cp}}(A) \leq \xi_t^{\mathrm{cp,isp}}(A) \leq \tau_{\mathrm{cp}}(A) \text{ for any } t \geq 1.$$

One may also define the sparse analogs of the constraints (36), (37), (38), and (39):

$$L_k([x(V_k)]_{2t}) \geq 0 \ \text{ for } k \in [p], \qquad (46)$$

$$L_k((\sqrt{A_{ii}}x_i - x_i^2)[x(V_k)]_{2t-2}) \geq 0 \ \text{ for } i \in V_k, \ k \in [p], \qquad (47)$$

$$L_k((A_{ij} - x_ix_j)[x(V_k)]_{2t-2}) \geq 0 \ \text{ for } \{i, j\} \subseteq V_k, \ k \in [p], \qquad (48)$$

$$L_k(x_ix_j[x(V_k)]_{t-1}[x(V_k)]_{t-1}^T) \succeq 0 \ \text{ for } i \neq j \in V_k, \ k \in [p]. \qquad (49)$$

Then, define $\xi_{t,\dagger}^{\mathrm{cp,isp}}(A)$ by adding constraint (48) to $\xi_t^{\mathrm{cp,isp}}(A)$, and $\xi_{t,\ddagger}^{\mathrm{cp,isp}}(A)$ by adding the constraints (46), (47) and (49) to $\xi_{t,\dagger}^{\mathrm{cp,isp}}(A)$, so that $\xi_t^{\mathrm{cp,isp}}(A) \leq \xi_{t,\dagger}^{\mathrm{cp,isp}}(A) \leq \xi_{t,\ddagger}^{\mathrm{cp,isp}}(A)$.

Weak ideal-sparse hierarchies for cp-rank. Observe that, if, in equation (45), we replace the matrix $A - xx^T$ by its principal submatrix indexed by V_k, then one also gets a lower bound on $\tau_{\mathrm{cp}}(A)$, possibly weaker than $\xi_t^{\mathrm{cp,isp}}(A)$, but potentially easier to compute. Let $\xi_t^{\mathrm{cp,wisp}}(A)$ denote the parameter obtained by replacing the condition (45) in the definition of $\xi_t^{\mathrm{cp,isp}}(A)$ by the following (weaker) constraint

$$L_k((A[V_k] - x(V_k)x(V_k)^T) \otimes [x(V_k)]_{t-1}[x(V_k)]_{t-1}^T) \succeq 0 \ \text{ for } k \in [p]. \qquad (50)$$

Then we have

$$\xi_t^{\mathrm{cp,wisp}}(A) \leq \xi_t^{\mathrm{cp,isp}}(A).$$

Since we have weakened some conditions of the ideal-sparse hierarchy, the weak ideal-sparse hierarchy $\xi_t^{\mathrm{cp,wisp}}(A)$ is no longer guaranteed to be at least as strong as the dense hierarchy $\xi_t^{\mathrm{cp}}(A)$. This is substantiated by our numerical experiments, where we frequently observe $\xi_t^{\mathrm{cp,wisp}}(A) < \xi_t^{\mathrm{cp}}(A)$ for randomly generated matrices A; see Sect. 4.3.1 for how we generate these matrices and see (53) for a concrete instance of such matrix. On the other hand, in all of the high cp-rank matrices A from the literature that we consider in Sect. 4.3.2, it does hold that $\xi_t^{\mathrm{cp}}(A) \leq \xi_t^{\mathrm{cp,wisp}}(A)$. This relation also holds for several other cp-rank matrices from the literature we have considered but did not present in this paper. It seems that the delineating factor might be that our randomly generated matrices tend to have cp-rank close to the usual matrix rank (i.e., $\mathrm{rank}_{\mathrm{cp}}(A) - \mathrm{rank}(A) \leq 1$), while, in contrast, the matrices considered in the literature have a cp-rank often much higher (e.g., up to 27 for Example ex4 in (55)) than the rank.

4.2 Links to combinatorial lower bounds on the cp-rank

We indicate here some links to other known lower bounds on the cp-rank. Clearly the rank is a lower bound:

$$\text{rank}_{\text{cp}}(A) \geq \text{rank}(A).$$

A combinatorial lower bound arises naturally from the edge clique-cover number of the support graph G_A.

Given a graph $G = (V, E)$, its *edge clique-cover number*, denoted $c(G)$ (following [29]), is defined as the smallest number of (maximal) cliques in G whose union covers every edge of G. This parameter is NP-hard to compute [30]. Clearly, $c(G) = |E|$ if G is a triangle-free graph (i.e., $\omega(G) = 2$, where $\omega(G)$ denotes the maximum cardinality of a clique in G). As observed in [29], the edge clique-cover parameter gives a lower bound on the cp-rank:

$$\text{rank}_{\text{cp}}(A) \geq c(G_A).$$

Indeed, if $A = \sum_{\ell=1}^{r} a_\ell a_\ell^T$ with $a_\ell \geq 0$ and $r = \text{rank}_{\text{cp}}(A)$, then the supports of $a_1, \ldots, a_r$ are (not necessarily distinct) cliques that provide an edge clique-cover of G_A by at most r cliques.

In [29] a semidefinite parameter $\tau_{\text{cp}}^{\text{sos}}(A)$ is introduced, which is shown to be at least as good as $\text{rank}(A)$ and as $c_{\text{frac}}(G_A)$, the *fractional edge clique-cover number*, i.e., the natural linear relaxation of $c(G_A)$ defined by

$$c_{\text{frac}}(G_A) = \min \left\{ \sum_{k=1}^{p} x_k : \sum_{k:\{i,j\} \subseteq V_k} x_k \geq 1 \text{ for } \{i, j\} \in E_A \right\}. \tag{51}$$

So, we have $c(G_A) \geq c_{\text{frac}}(G_A)$ and

$$\tau_{\text{cp}}(A) \geq \tau_{\text{cp}}^{\text{sos}}(A) \geq \max\{\text{rank}(A), c_{\text{frac}}(G_A)\}.$$

In [33] it is shown that the bound $\xi_{2,(2019),\dagger}^{\text{cp}}(A)$ is at least as strong as $\tau_{\text{cp}}^{\text{sos}}(A)$. Indeed, the proof for the relevant result (Proposition 7 in [33]) only uses the relation $L((A_{ij} - x_i x_j) x_i x_j) \geq 0$ from (38) and the relation $L((xx^T)^{\otimes 2}) \preceq A^{\otimes 2}$ in (35). Hence, we have the chain of inequalities

$$\tau_{\text{cp}}(A) \geq \xi_{2,\dagger}^{\text{cp,isp}}(A) \geq \xi_{2,\dagger}^{\text{cp}}(A) \geq \xi_{2,(2019),\dagger}^{\text{cp}}(A) \geq \tau_{\text{cp}}^{\text{sos}}(A)$$
$$\geq \max\{\text{rank}(A), c_{\text{frac}}(G_A)\}. \tag{52}$$

As we now observe, the (weak) ideal-sparse bound $\xi_1^{\text{cp,wisp}}(A)$ of the first level $t = 1$ is at least as good as the parameter $c_{\text{frac}}(G_A)$.

Lemma 13 *If $A \in \mathcal{S}^n$ is nonnegative with support graph G_A, then $\xi_1^{\text{cp,wisp}}(A) \geq c_{\text{frac}}(G_A)$.*

Proof Let $(L_1, \ldots, L_p)$ be an optimal solution for the parameter $\xi_1^{\mathrm{cp,wisp}}(A)$. Using (44), we have

$$L_k(A_{ij} - x_i x_j) \geq 0 \text{ for all } i \neq j \text{ with } \{i, j\} \subseteq V_k \text{ and } k \in [p],$$

which gives $A_{ij} L_k(1) \geq L_k(x_i x_j)$. Summing over k, we get

$$A_{ij} \sum_{k \in [p]:\{i,j\}\subseteq V_k} L_k(1) \geq \sum_{k \in [p]:\{i,j\}\subseteq V_k} L_k(x_i x_j) = A_{ij},$$

where we use (41) for the last equality. As $A_{ij} > 0$, this gives $\sum_{k:\{i,j\}\subseteq V_k} L_k(1) \geq 1$ for every edge $\{i, j\} \in E_A$. Hence, the vector $x = (L_k(1))_{k=1}^p \in \mathbb{R}_+^p$ is feasible for program (51), which implies the inequality $\sum_{k=1}^p L_k(1) \geq c_{\mathrm{frac}}(G_A)$, as desired. $\square$

We now give a class of cp-matrices that exhibit a large separation between the dense and ideal-sparse bounds at level $t = 1$: these matrices have size $n = 2m$ and $\mathrm{rank}_{\mathrm{cp}}(A) = \xi_1^{\mathrm{cp,wisp}}(A) = m^2 \geq m + 1 > \xi_1^{\mathrm{cp}}(A)$.

Example 14 For $n = 2m$ consider the matrix

$$A = \begin{pmatrix} (m+1)I_m & J_m \\ J_m & (m+1)I_m \end{pmatrix} \in \mathcal{S}^n,$$

where I_m is the identity matrix and J_m the all-ones matrix. Then, A is a cp-matrix (because it is nonnegative and diagonally dominant). Its cp-rank is $\mathrm{rank}_{\mathrm{cp}}(A) = |E_A| = m^2$ (because its support graph G_A is the complete bipartite graph $K_{m,m}$ (thus, connected, triangle-free and not a tree), using a result of [26], also mentioned below). Clearly, we have $c(G_A) = c_{\mathrm{frac}}(G_A) = m^2$. Hence, using Lemma 13, we obtain $\xi_1^{\mathrm{cp,isp}}(A) = \xi_1^{\mathrm{cp,wisp}}(A) = m^2 = \mathrm{rank}_{\mathrm{cp}}(A)$. We claim $\xi_1^{\mathrm{cp}}(A) < m + 1$, which shows a large separation between the dense and ideal-sparse bounds of level $t = 1$.

For this, observe that $\xi_1^{\mathrm{cp}}(A)$ can be reformulated as

$$\xi_1^{\mathrm{cp}}(A) = \min\{L(1) : L \in \mathbb{R}[x]_2^*, \ L(1) \geq 1, \ L(x_i) \geq \sqrt{A_{ii}} \ (i \in [n]),$$
$$L(xx^T) = A, \ L([x]_1[x]_1^T) \succeq 0\}.$$

Consider the linear functional $L \in \mathbb{R}[x]_2^*$ defined by $L(xx^T) = A$, $L(x_i) = \sqrt{m+1}$ for $i \in [n]$ and $L(1) = \frac{2m(m+1)}{2m+1}$. We show that L is feasible for the above program, which implies $\xi_1^{\mathrm{cp}}(A) \leq L(1) < m + 1$. For this, it suffices to show $L([x]_1[x]_1^T) \succeq 0$. By taking the Schur complement with respect to the upper left corner, this boils down to checking that $L(1)A - (m+1)J_m \succeq 0$. As the all-ones vector is an eigenvector of A (with eigenvalue $2m + 1$), it is also an eigenvector of $L(1)A - (m+1)J_m$ with corresponding eigenvalue $L(1)(2m + 1) - 2m(m + 1) = 0$. The eigenvalues of the matrix $L(1)A - (m+1)J_m$ for its eigenvectors orthogonal to the all-ones vector are eigenvalues of A, and thus they are nonnegative since A is positive semidefinite. This shows that $L(1)A - (m + 1)J_m \succeq 0$ and the proof is complete.

We conclude with some observations about known upper bounds on the cp-rank.

General upper bounds on the cp-rank are $\text{rank}_{cp}(A) \leq n$ if $n \leq 4$, $\text{rank}_{cp}(A) \leq \binom{n+1}{2} - 4$ if $n \geq 5$ [58], and $\text{rank}_{cp}(A) \leq \binom{r+1}{2} - 1$ if $r = \text{rank}(A) \geq 2$ [7].

It is known that $c(G) \leq n^2/4$ [28]. In analogy, it has been a long standing conjecture by Drew et al. [26] that the cp-rank of an $n \times n$ completely positive matrix is at most $n^2/4$. This conjecture, however, was disproved in [11, 12] for any $n \geq 7$. In particular, it is shown in [12] that the maximum cp-rank of an $n \times n$ cp-matrix is of the order $n^2/2 + O(n^{3/2})$.

If the support graph G_A is triangle-free, then $|E_A| \leq \text{rank}_{cp}(A) \leq \max\{n, |E_A|\}$; moreover, if G_A is connected, triangle-free and not a tree, then $\text{rank}_{cp}(A) = |E_A|$ [26]. Hence, $n - 1 = |E_A| \leq \text{rank}_{cp}(A) \leq n$ if G_A is a tree, with $\text{rank}_{cp}(A) = n$ if A is nonsingular. By Lemma 13, we know that $\xi_1^{cp,\text{wisp}}(A) \geq |E_A|$ if G_A is triangle-free. Hence, the bound $\xi_1^{cp,\text{wisp}}(A)$ gives the exact value of the cp-rank when G_A is connected, triangle-free and not a tree. On the other hand, if G_A is a tree and A is nonsingular, then the bound $\xi_{2,\dagger}^{cp}(A)$ gives the exact value (equal to n) of the cp-rank (since it is at least $\tau_{cp}^{sos}(A) \geq \text{rank}(A)$ by relation (52)).

4.3 Numerical results for the completely positive rank

In this section, we explore the behaviour of the various bounds for the completely positive rank on three classes of examples. Our objective is to illustrate the superiority of the ideal-sparse hierarchies compared to the dense ones. We examine both the quality of the bounds as well as computation times.

The first class we consider consists of randomly generated sparse cp-matrices. We will give the exact construction below. In all numerical examples we considered for these matrices, the bounds obtained for $\xi_t^{cp}(A)$ and $\xi_t^{cp,\text{isp}}(A)$ were always at most rank$(A)+2$. So we do not list the numerical bounds for these examples as there does not seem to be much insight gained from them. However, random examples give us a way to compare the computation times amongst different hierarchies and across various matrix sizes, non-zero densities, and levels. In what follows the *non-zero density* of $A \in \mathcal{S}^n$, denoted nzd(A), is defined as the proportion of non-zero entries above the main diagonal, i.e., nzd$(A) = |E_A|/\binom{n}{2}$. Hence, a diagonal matrix has nzd=0, and a dense matrix has nzd=1.

The second class contains examples from the literature, whose cp-rank is known from theory. However, recall the moment hierarchies provide lower bounds on τ_{cp}, whose value is often unknown and could be strictly less than the cp-rank. Regardless, these examples give an interesting testbed to evaluate the quality of the new bounds.

The third class of examples consists of doubly nonnegative matrices, which are known to not be completely positive. In running these examples, the hope is to obtain an infeasibility certificate from the solver. This then numerically certifies that the matrix is not completely positive. In this context one hierarchy is said to perform better than another one if it returns the infeasibility certificate at a lower level or using less run time.
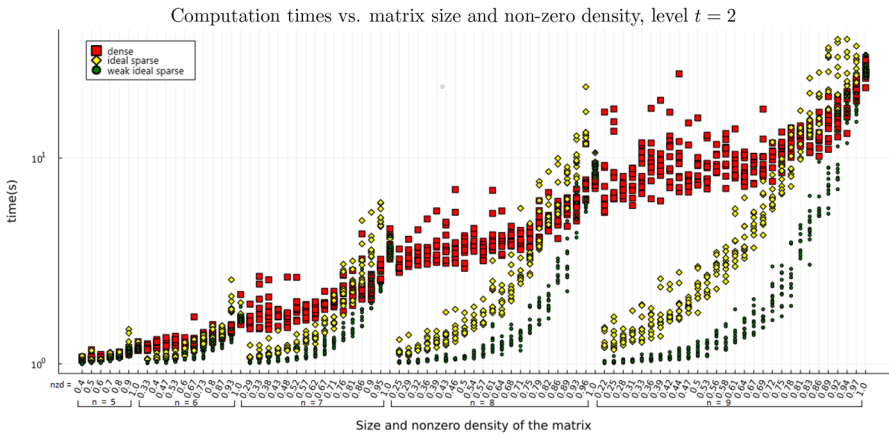
Fig. 1 Scatter plot of the computation times (in seconds) for the three hierarchies $\xi_{2,\dagger}^{\mathrm{cp}}$ (indicated by a red square), $\xi_{2,\dagger}^{\mathrm{cp,isp}}$ (indicated by a yellow losange), $\xi_{2,\dagger}^{\mathrm{cp,wisp}}$ (indicated by a green circle) against matrix size and non-zero density for 850 random matrices, generated using the above described procedure. The matrices are arranged in ascending size ($n = 5, 6, 7, 8, 9$) and then ascending non-zero density, ranging from the minimal density needed to have a connected support graph up to a fully dense matrix (nzd = 1)

The size of the matrices involved in the semidefinite programs grows quickly with the level t in the hierarchy (roughly, as $\binom{n+t}{t}$), so these problems become quickly too big for the solver (in particular, due to memory storage). We will consider matrices up to size $n = 12$ for the dense and sparse hierarchies at level $t = 2$. At level $t = 3$ and for matrices of size $n = 12$, we can only compute bounds for the weak sparse hierarchy.

All computations shown were run on a personal computer running Windows 11 Home 64-bit with an 11th Gen Intel(R) Core(TM) i7-11800 H @ 2.30GHz Processor and 16GB of RAM. The software we use was custom coded in Julia [9] utilizing the JuMP [27] package for problem formulation, and MOSEK [3] as the semidefinite programming solver.[2]

4.3.1 Randomly generated sparse cp-matrices

We first describe how we construct random sparse cp-matrices. Given integers $n \in \mathbb{N}$ and $n - 1 \leq m \leq \binom{n}{2}$, we create a symmetric $n \times n$ binary matrix M with exactly m ones above the diagonal, whose positions are selected uniformly at random. Let G be the graph with M as adjacency matrix. We only keep the instances where G is a connected graph. We enumerate the maximal cliques $V_1, \ldots, V_p$ of G (using, e.g., the Bron-Kerbosch algorithm [14]). Then, we select a subset of maximal cliques $V_{q_1}, \ldots, V_{q_l}$ whose union covers every edge of G (e.g., using a greedy algorithm). For each $k \in [l]$, generate $m_k \geq 1$ vectors $(a^{(k,i)})_{i \in [m_k]} \subseteq \mathbb{R}_+^n$ with uniformly random entries following $\mathcal{U}[0, 1]$ and supported on V_{q_k}. We will choose $m_k = 2$ by default. Then, consider the matrix $\sum_{k \in [l]} \sum_{i \in [m_k]} a^{(k,i)}(a^{(k,i)})^T$, scale it so that all diagonal
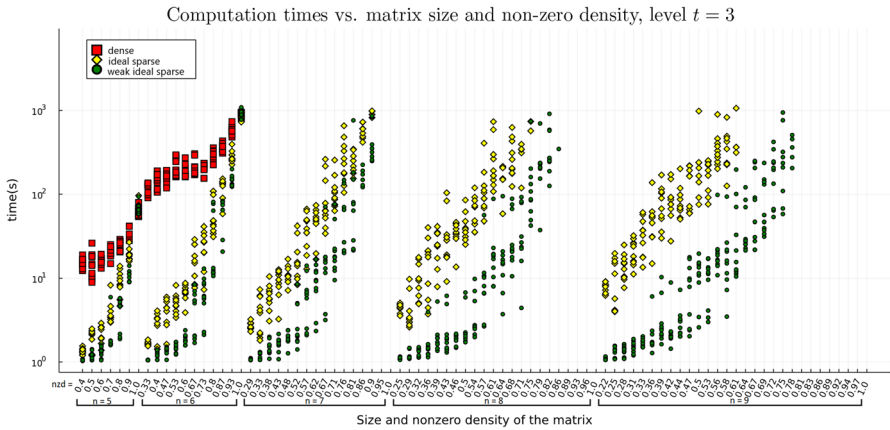

Fig. 2 This is a similar plot to Fig. 1 but now for level t=3 of each of the hierarchies. By omitting markers we indicate that the corresponding computations either exceeded memory constraints or took longer than 10^3 seconds

entries are equal to 1 and call A the resulting matrix. By construction, A is completely positive with connected support $G_A = G$, and non-zero density nzd $= m/\binom{n}{2}$.

We generate such random examples for varying matrix size ($n = 5, 6, 7, 8, 9$) and incrementing the non-zero density nzd in ascending order. In order to not include examples with disconnected graphs we need nzd $\geq (n-1)/\binom{n}{2}$. To account for different graph configurations with the same non-zero density we generate 10 examples per matrix size and nzd value. For all of them we compute the dense and (weak) sparse bounds of level $t = 2$ and $t = 3$. Here, we are not so much interested in the numerical bounds, but rather in their computation times. This numerical experiment indeed permits to show the differences in computation time between the ideal-sparse and dense hierarchies. It turns out that the computation times for the parameters ξ_t^{cp}, $\xi_{t,\dagger}^{cp}$, and $\xi_{t,\ddagger}^{cp}$ are all comparable at level $t = 2, 3$, likewise for the ideal-sparse analogs. For this reason, we only plot the results for the "†" variant, i.e., for the parameters $\xi_{t,\dagger}^{cp}$, $\xi_{t,\dagger}^{cp,isp}$, $\xi_{t,\dagger}^{cp,wisp}$. The results are shown in Fig. 1 (for $t = 2$) and in Fig. 2 (for $t = 3$).

We can make the following observations about the results in Fig. 1. As expected, the ideal-sparse hierarchy is faster to compute than the dense hierarchy for matrices with non-zero density nzd ≤ 0.8. The computation of the weak ideal-sparse hierarchy is even faster. Moreover, the speed-up increases with the matrix size and the level of the hierarchy as can be seen across Figs. 1 and 2. At level $t = 3$, some hierarchies can no longer be computed for certain matrix sizes and non-zero densities. This is particularly evident in the case of the dense hierarchy for matrices of size 7 and larger. The ideal-sparse hierarchies can be computed up to size 9 depending on the non-zero density. We show only the examples that we could compute in less than 10^3 seconds. The parameters that either took longer that 10^3 seconds or exceeded memory constraints can be inferred by the omission of their respective markers in Fig. 2.

We also make an observation regarding how the values of the dense and weak-ideal sparse bounds compare for these random matrices. As observed earlier, the weak ideal-sparse hierarchy $\xi_t^{cp,wisp}(A)$ is no longer guaranteed to be at least as strong as the dense hierarchy $\xi_t^{cp}(A)$. Indeed, in our numerical experiments, we frequently observe

the strict inequality $\xi_t^{\mathrm{cp,wisp}}(A) < \xi_t^{\mathrm{cp}}(A)$ for randomly generated matrices A. For example, the matrix (with entries rounded for presentation)

$$A = \begin{pmatrix} 1.0 & 0.578 & 0.0 & 0.0 & 0.225 \\ 0.578 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.656 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.526 \\ 0.225 & 0.0 & 0.656 & 0.526 & 1.0 \end{pmatrix} \tag{53}$$

has the following parameters at order $t = 2$:

$$\left(\xi_2^{\mathrm{cp,wisp}}(A) = 4\right) < \left(\xi_2^{\mathrm{cp}}(A) = 5\right) \le \left(\xi_2^{\mathrm{cp,isp}}(A) = 5\right) \le \left(\mathrm{rank}_{\mathrm{cp}}(A) = 5\right).$$

4.3.2 Selected sparse cp-matrices

Here, we compute the dense and (weak) ideal-sparse parameters for a few selected cp-matrices taken from the literature. We first briefly discuss the four example matrices we will consider, denoted ex1, ex2, ex3, ex4, and shown in relations (54) and (55) below.

$$\mathrm{ex1} = \begin{pmatrix} 3 & 2 & 0 & 0 & 1 \\ 2 & 5 & 6 & 0 & 0 \\ 0 & 6 & 14 & 4 & 0 \\ 0 & 0 & 4 & 9 & 1 \\ 1 & 0 & 0 & 1 & 2 \end{pmatrix}, \quad \mathrm{ex3} = \begin{pmatrix} 781 & 0 & 72 & 36 & 228 & 320 & 240 & 228 & 36 & 96 & 0 \\ 0 & 845 & 0 & 96 & 36 & 228 & 320 & 320 & 228 & 36 & 96 \\ 72 & 0 & 827 & 0 & 72 & 36 & 198 & 320 & 320 & 198 & 36 \\ 36 & 96 & 0 & 845 & 0 & 96 & 36 & 228 & 320 & 320 & 228 \\ 228 & 36 & 72 & 0 & 781 & 0 & 96 & 36 & 228 & 240 & 320 \\ 320 & 228 & 36 & 96 & 0 & 845 & 0 & 96 & 36 & 228 & 320 \\ 240 & 320 & 198 & 36 & 96 & 0 & 745 & 0 & 96 & 36 & 228 \\ 228 & 320 & 320 & 228 & 36 & 96 & 0 & 845 & 0 & 96 & 36 \\ 36 & 228 & 320 & 320 & 228 & 36 & 96 & 0 & 845 & 0 & 96 \\ 96 & 36 & 198 & 320 & 240 & 228 & 36 & 96 & 0 & 745 & 0 \\ 0 & 96 & 36 & 228 & 320 & 320 & 228 & 36 & 96 & 0 & 845 \end{pmatrix}, \tag{54}$$

$$\mathrm{ex2} = \begin{pmatrix} 2 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 3 & 0 \\ 1 & 1 & 1 & 0 & 3 \end{pmatrix}, \quad \mathrm{ex4} = \begin{pmatrix} 91 & 0 & 0 & 0 & 19 & 24 & 24 & 24 & 19 & 24 & 24 & 24 \\ 0 & 42 & 0 & 0 & 24 & 6 & 6 & 6 & 24 & 6 & 6 & 6 \\ 0 & 0 & 42 & 0 & 24 & 6 & 6 & 6 & 24 & 6 & 6 & 6 \\ 0 & 0 & 0 & 42 & 24 & 6 & 6 & 6 & 24 & 6 & 6 & 6 \\ 19 & 24 & 24 & 24 & 91 & 0 & 0 & 0 & 19 & 24 & 24 & 24 \\ 24 & 6 & 6 & 6 & 0 & 42 & 0 & 0 & 24 & 6 & 6 & 6 \\ 24 & 6 & 6 & 6 & 0 & 0 & 42 & 0 & 24 & 6 & 6 & 6 \\ 24 & 6 & 6 & 6 & 0 & 0 & 0 & 42 & 24 & 6 & 6 & 6 \\ 19 & 24 & 24 & 24 & 19 & 24 & 24 & 24 & 91 & 0 & 0 & 0 \\ 24 & 6 & 6 & 6 & 24 & 6 & 6 & 6 & 0 & 42 & 0 & 0 \\ 24 & 6 & 6 & 6 & 24 & 6 & 6 & 6 & 0 & 0 & 42 & 0 \\ 24 & 6 & 6 & 6 & 24 & 6 & 6 & 6 & 0 & 0 & 0 & 42 \end{pmatrix}. \tag{55}$$

Table 1 Dense and ideal-sparse bounds for selected sparse cp-matrices

A	n	p	c	r	Bounds			r_{cp}	Times (s)		
					ξ_1^{cp}	$\xi_1^{cp,isp}$	$\xi_1^{cp,wisp}$		Dense	Ideal-sparse	wisp
ex1	5	5	5	5	2.71	5	5	5	< 1	< 1	< 1
ex2	5	6	6	4	3	6	6	6	< 1	< 1	< 1
ex3	11	22	8	11	4.24	8.53	8.53	32	< 1	< 1	< 1
ex4	12	64	16	10	4.85	29.66	29.63	37	< 1	< 1	< 1
					$\xi_{2,\ddagger}^{cp}$	$\xi_{2,\ddagger}^{cp,isp}$	$\xi_{2,\ddagger}^{cp,wisp}$				
ex1	5			5	5	5	5	5	< 1	< 1	< 1
ex2	5			4	6	6	6	6	< 1	< 1	< 1
ex3	11			11	21.93	22.32	22.32	32	123.86	54.89	8.14
ex4	12			10	29.57	29.66	29.66	37	238.94	33.78	1.28
					$\xi_{3,\ddagger}^{cp}$	$\xi_{3,\ddagger}^{cp,isp}$	$\xi_{3,\ddagger}^{cp,wisp}$				
ex3	11			11	–	–	22.33	32	–	–	2648.69
ex4	12			10	–	–	29.66	37	–	–	28.69

n = size of A, p = number of maximal cliques of G_A, c = edge clique-cover number of G_A, r = rank(A), r_{cp} = rank$_{cp}(A)$

– : computations failed due to memory constraints

The matrix ex1 (from [6]) is supported on the 5-cycle C_5 and the matrix ex2 (from [68]) is supported on the bipartite graph $K_{3,2}$. In both cases, we have $\xi_1^{cp,isp}(A) =$ rank$_{cp}(A) = |E_A|$ (combining Lemma 13 and the results of [26] mentioned earlier at the end of Sect. 4.2). The matrices ex3 and ex4 were constructed, respectively, in [11, 12] as examples of matrices having a large cp-rank exceeding the value $n^2/4$ (thus refuting the conjecture by Drew et al. [26]). The matrix ex3 is supported on $\overline{C_{11}}$, the complement of an 11-cycle, and matrix ex4 is supported on the complete tripartite graph $K_{4,4,4}$. One can verify that the edge clique-cover number is equal to 8 for $\overline{C_{11}}$ and to 16 for $K_{4,4,4}$.

The numerical results for these four examples are presented in Table 1, where we also show other parameters for the matrix (size n, rank r, cp-rank r_{cp}) and its support graph (number p of maximal cliques, edge clique-cover number c). Here are some comments about Table 1.

The results confirm the results in Lemma 13: the ideal-sparse bound of level $t = 1$ is equal to the number of edges for ex1 and ex2 (and matches the cp-rank); moreover it gives a strong improvement on the dense bound of level 1. The bounds of level $t = 2$ all exceed the rank of the matrix (as expected in view of (52)). At level $t = 3$, only the weak ideal-sparse bound can be computed for the matrices ex3 and ex4.

In Table 1, the values of the bounds at level $t = 3$ are close to those at level $t = 2$ for matrices ex3 and ex4. However, the tests for the flatness condition (21) fail, so that one cannot claim that the bounds are equal to τ_{cp} at this stage.

We also tested whether the flatness conditions (20) and (21) hold for matrices ex1 and ex2 at level $t = 2$, and whether one can extract atoms and construct a cp-factorization.

Table 2 Testing flatness and atom extraction

A	$\xi_{2,\ddagger}^{cp}$		$\xi_{2,\ddagger}^{cp,isp}$		$\xi_{2,\ddagger}^{cp,wisp}$	
	Flatness (20)	# Atoms	Flatness (21)	# Atoms	Flatness (21)	# Atoms
ex1	False	0	True	10	False	0
ex2	False	6	True	6	True	6
	$\xi_{3,\ddagger}^{cp}$		$\xi_{3,\ddagger}^{cp,isp}$		$\xi_{3,\ddagger}^{cp,wisp}$	
ex1	False	10	True	10	False	0
ex2	True	6	True	6	True	6

The results are summarized in Table 2, where we indicate the number of atoms (corresponding to a cp-factorization with that many factors) when the extraction procedure is successful. We indicate that the extraction procedure fails by reporting "# atoms=0". As mentioned in [37], one may indeed try and apply the extraction procedure even if flatness does not hold.

For the dense bounds of level $t = 2$, flatness does not hold for the matrices ex1 and ex2. However, while one does not succeed to extract atoms for matrix ex1, the extraction is successful for matrix ex2 and returns 6 atoms. Interestingly, flatness holds for the ideal-sparse bounds and the atom extraction is successful. However, the number of extracted atoms is 10 for matrix ex1, thus twice the cp-rank. To verify that the extracted atoms are (approximatively) correct, we use them to construct a cp-matrix A_{rec}, which we then compare to the original matrix A. In all cases we obtain $\|A_{\mathrm{rec}} - A\|_1 \leq 10^{-8}$, which shows that a correct factorization has been constructed.

Note that for the ideal-sparse parameter, since one splits the problem over the maximal cliques and has a distinct linear functional L_k for each clique V_k, it may be more difficult to satisfy the flatness condition (21) (since each L_k must satisfy it), as happens for matrices ex3 and ex4.

4.3.3 Doubly nonnegative matrices that are not completely positive

In this section we consider the following three matrices that are known to be doubly nonnegative but not completely positive (taken from [6, 53, 57]):

$$\text{ex5} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 1 & 0 & 0 & 1 & 3 \end{pmatrix}, \quad \text{ex6} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 1 & 0 & 0 & 1 & 6 \end{pmatrix}, \quad \text{ex7} = \begin{pmatrix} 7 & 1 & 2 & 2 & 1 & 1 \\ 1 & 12 & 1 & 3 & 3 & 5 \\ 2 & 1 & 2 & 3 & 0 & 0 \\ 2 & 3 & 3 & 5 & 0 & 0 \\ 1 & 3 & 0 & 0 & 2 & 4 \\ 1 & 5 & 0 & 0 & 4 & 10 \end{pmatrix}.$$

The objective is to see whether the hierarchies are able to detect that the matrix is not cp. This can be achieved in two ways: when the solver returns an infeasibility certificate, or when it returns a bound that exceeds a known upper bound on the cp-rank. We test this for the bounds at level $t = 1$ and $t = 2$. At level $t = 2$ we try

Table 3 Detecting non-cp matrices for $t = 1$

A	n	r	ξ_1^{cp}	$\xi_1^{cp,isp}$	$\xi_1^{cp,wisp}$	$r_{cp} \leq$
ex5	5	4	2.47	*	*	5
ex6	5	5	2.59	*	*	5
ex7	6	6	2.4	3.02	3.02	17

* = infeasibility certificate

different variants by adding the constraints (36), (37), (38), and (39) and their sparse analogs. The results are presented in Tables 3 and 4.

There we indicate one of three possible outcomes. The first outcome is indicated with a question mark "?", which indicates that the solver could not reach a decision within the default MOSEK solver parameters. The second possible outcome is when the solver returns an infeasibility certificate (indicated with *), or when it returns a value that exceeds a known upper bound for the cp-rank (in which case the bound is marked with *). The last column in both tables, labeled "$r_{cp} \leq$", provides such an upper bound on the cp-rank of a cp-matrix with the given support graph. The third possible outcome is when the solver returns a value that does not violate the upper bound, in which case no conclusion can be reached. All computations took less than a second and hence times are not shown.

We make three observations about Tables 3 and 4. The first is that the ideal-sparse hierarchies show infeasibility at level $t = 1$ already for examples ex5 and ex6 while the dense hierarchy shows the same only at level $t = 2$ with all additional constraints imposed. Secondly, the ideal-sparse hierarchy correctly identifies ex7 as not cp at level $t = 2$ while the dense hierarchy does not succeed even at level $t = 3$. The third observation is that adding additional constraints helps prevent the solver from returning an "unknown result status" but this seems to be less needed in the case of the ideal-sparse hierarchies. It should be noted that increasing the level of the hierarchy creates more opportunity for numerical errors in the computations, as seen in Table 4.

5 Application to the nonnegative rank

In this section we indicate how the treatment in the previous section for the cp-rank extends naturally to the asymmetric setting of the nonnegative rank.

5.1 Ideal-sparsity bounds for the nonnegative rank

Given a nonnegative matrix $M \in \mathbb{R}^{m \times n}$, its *nonnegative rank*, denoted $\text{rank}_+(M)$, is the smallest integer r for which there exist nonnegative vectors $a_\ell \in \mathbb{R}_+^m$ and $b_\ell \in \mathbb{R}_+^n$ such that

$$M = \sum_{\ell=1}^{r} a_\ell b_\ell^T. \tag{56}$$

Table 4 Detecting non cp-matrices for $t = 2, 3$

A	n	r	ξ_2^{cp}	$\xi_2^{cp,isp}$	$\xi_2^{cp,wisp}$	$\xi_{2,\dagger}^{cp}$	$\xi_{2,\dagger}^{cp,isp}$	$\xi_{2,\dagger}^{cp,wisp}$	$\xi_{2,\ddagger}^{cp}$	$\xi_{2,\ddagger}^{cp,isp}$	$\xi_{2,\ddagger}^{cp,wisp}$	$r_{cp} \leq$
ex5	5	4	?	*	?	?	*	*	*	*	*	5
ex6	5	5	13.56*	?	*	13.56*	*	*	16.11*	*	*	5
ex7	6	6	?	34.88*	34.01*	12.94	*	*	13.89	*	*	17

A	n	r	ξ_3^{cp}	$\xi_3^{cp,isp}$	$\xi_3^{cp,wisp}$	$\xi_{3,\dagger}^{cp}$	$\xi_{3,\dagger}^{cp,isp}$	$\xi_{3,\dagger}^{cp,wisp}$	$\xi_{3,\ddagger}^{cp}$	$\xi_{3,\ddagger}^{cp,isp}$	$\xi_{3,\ddagger}^{cp,wisp}$	$r_{cp} \leq$
ex5	5	4	?	*	?	*	*	*	*	*	*	5
ex6	5	5	?	*	*	?	*	*	194.19*	*	*	5
ex7	6	6	?	?	?	?	*	*	?	*	*	17

$* =$ infeasibility certificate, $? =$ unknown result status

Computing the nonnegative rank is an NP-hard problem [62]. Fawzi and Parrilo [29] introduced the following natural "convexification" of the nonnegative rank:

$$\tau_+(M) = \inf\left\{\lambda : \frac{1}{\lambda}M \in \mathrm{conv}\{xy^T : x \in \mathbb{R}^m_+, y \in \mathbb{R}^n_+, M \geq xy^T\}\right\},$$

which can be seen as an asymmetric analog of τ_{cp}. We consider the analogs of the parameters ξ_t^{cp} and $\xi_t^{cp,isp}$, which now involve linear functionals acting on polynomials in $m+n$ variables. As in the introduction, set $V = [m+n] = U \cup W$, where $U = [m] = \{1, \ldots, m\}$ (corresponding to the row indices of M) and $W = \{m + 1, \ldots, m + n\}$ (corresponding to the column indices of M, up to a shift by m). Set

$$E^M = \{\{i, j\} \in U \times W : M_{i, j-m} \neq 0\},$$

so that the bipartite graph $G^M = (V = U \cup W, E^M)$ corresponds to the support graph of M. We also set $\overline{E}^M = (U \times W)\backslash E^M$ and $M_{\max} = \max_{i \in U, j \in W} M_{i, j-m}$. As is well-known (see, e.g., [33]), the vectors in (56) may be assumed to satisfy $\|a_\ell\|_\infty, \|b_\ell\|_\infty \leq \sqrt{M_{\max}}$ (after rescaling). This motivates the definition of the semialgebraic set K^M from (19) and, for any integer $t \geq 1$, of the parameter:

$$\xi_t^+(M) = \min\{L(1) : L \in \mathbb{R}[x_1, \ldots, x_{m+n}]_{2t}^*, \tag{57}$$
$$L(x_i x_j) = M_{i, j-m} \ (i \in U, j \in W), \tag{58}$$
$$L([x]_t[x]_t^T) \succeq 0, \tag{59}$$
$$L((\sqrt{M_{\max}}x_i - x_i^2)[x]_{t-1}[x]_{t-1}^T) \succeq 0 \ \text{for } i \in V, \tag{60}$$
$$L((M_{i, j-m} - x_i x_j)[x]_{t-1}[x]_{t-1}^T) \succeq 0 \ \text{for } \{i, j\} \in E^M, \tag{61}$$
$$L(x_i x_j[x]_{2t-2}) = 0 \ \text{for } \{i, j\} \in \overline{E}^M\}. \tag{62}$$

If we omit the (ideal) constraint (62) and require the constraint (61) to hold also for pairs $\{i, j\} \in \overline{E}^M$, then we obtain the (weaker) parameter $\xi_t^+(M)$, introduced in [33] as a lower bound on $\tau_+(M)$ (and thus on $\mathrm{rank}_+(M)$).

In addition, we can define ideal-sparse bounds, by further exploiting the sparsity pattern of M. As the support graph G^M is now a bipartite graph it is convenient to use the following notion of biclique. A *biclique* in G^M corresponds to a complete bipartite subgraph and it is thus given by a pair (A, B) with $A \subseteq U$ and $B \subseteq W$ such that $\{i, j\} \in E^M$ for all $(i, j) \in A \times B$; it is maximal if $A \cup B$ is maximal. Let

$V_1 = A_1 \cup B_1, \ldots, V_p = A_p \cup B_p$ be the vertex sets of the maximal bicliques in G^M and, for any integer $t \geq_p 1$, define the parameter

$$\xi_t^{+,\mathrm{isp}}(M) = \min \left\{ \sum_{k=1} L_k(1) : L_k \in \mathbb{R}[x(V_k)]_{2t}^* \ (k \in [p]), \right. \tag{63}$$

$$\sum_{k \in [p]:\{i,j\} \subseteq V_k} L_k(x_i x_j) = M_{i,j-m} \ (i \in U, j \in W), \tag{64}$$

$$L_k([x(V_k)]_t [x(V_k)]_t^T) \succeq 0 \ (k \in [p]), \tag{65}$$

$$L_k((\sqrt{M_{\max}}x_i - x_i^2)[x(V_k)]_{t-1}[x(V_k)]_{t-1}^T) \succeq 0 \ (i \in V_k, \ k \in [p]), \tag{66}$$

$$L_k((M_{i,j-m} - x_i x_j)[x(V_k)]_{t-1}[x(V_k)]_{t-1}^T) \succeq 0 \ (i \in U, j \in W,$$

$$\left. \{i, j\} \subseteq V_k, \ k \in [p]) \right\}. \tag{67}$$

Summarizing, we have the following inequalities among the above parameters

$$\xi_{t-1}^+(M) \leq \xi_t^+(M) \leq \xi_t^{+,\mathrm{isp}}(M) \leq \tau_+(M) \leq \mathrm{rank}_+(M) \ \text{ for any } t \geq 2,$$

with asymptotic convergence of all bounds to $\tau_+(M)$; this was shown in [33] for the bounds $\xi_t^+(M)$ (and this also follows as an application of Theorem 1).

As in the case of the cp-rank, there are more constraints that may be added to the above programs to strengthen the bounds. In [33] the authors propose to exploit the nonnegativity of the variables and add the constraints

$$L([x]_{2t}) \geq 0, \tag{68}$$

$$L((\sqrt{M_{\max}}x_i - x_i^2)[x]_{2t-2}) \geq 0 \text{ for } i \in V, \tag{69}$$

$$L((M_{i,j-m} - x_i x_j)[x]_{2t-2}) \geq 0 \text{ for } (i, j) \in U \times W. \tag{70}$$

Let $\xi_{t,\dagger}^+(M)$ denote the parameter obtained by adding the constraint (70) to $\xi_t^+(M)$. Similarly, one may add (70) to the parameter $\xi_t^+(M)$ (requiring (70) only for pairs in E^M) and its sparse analog to $\xi_t^{+,\mathrm{isp}}(M)$, leading, respectively, to the parameters $\xi_{t,\dagger}^+(M)$ and $\xi_{t,\dagger}^{+,\mathrm{isp}}(M)$. So, $\xi_{t,\dagger}^+(M) \leq \xi_{t,\dagger}^+(M) \leq \xi_{t,\dagger}^{+,\mathrm{isp}}(M)$. Finally, we also introduce the parameters, where we use the symbol ‡ instead of † when adding all the constraints (68), (69), (70).

5.2 Links to combinatorial lower bounds on the nonnegative rank

We now recall some other known lower bounds on the nonnegative rank and indicate their relations to the parameters considered here.

Fawzi and Parrilo [29] introduced a semidefinite bound $\tau_+^{\mathrm{sos}}(M)$ and show it satisfies $\tau_+^{\mathrm{sos}}(M) \leq \tau_+(M)$. In [33] it is shown that the parameters $\xi_{2,\dagger}^+(M)$ strengthen this bound[3]:

---

[3] This follows from the proof of [33, Proposition 15] since it only uses the relation $L((M_{i,j-m} - x_i x_j)x_i x_j) \geq 0$ for any $(i, j) \in U \times W$ in addition to the constraints defining the basic parameter $\xi_2^+(M)$.

$$\tau_+^{\text{sos}}(M) \le \xi_{2,\dagger}^+(M) \le \tau_+(M).$$

There is a well-known combinatorial lower bound on the nonnegative rank, which can be seen as an asymmetric analog of the lower bound on the cp-rank of A given by the edge clique-cover number $c(G_A)$. Recall $G^M = (U \cup W, E^M)$ is the bipartite graph defined as the support graph of $M \in \mathbb{R}_+^{m \times n}$. Define the *edge biclique-cover number* of G^M, denoted $\text{bc}(G^M)$, as the smallest number of bicliques whose union covers every edge in E^M. Then, we have

$$\text{rank}_+(M) \ge \text{bc}(G^M).$$

As a biclique in G^M corresponds to a pair $(A, B) \subseteq U \times W$ for which the rectangle $A \times B$ is fully contained in the support of M, the parameter $\text{bc}(G^M)$ is also known as the *rectangle covering number* of M (see, e.g., [29, 32]). Define its fractional analog $\text{bc}_{\text{frac}}(G^M)$ as

$$\text{bc}_{\text{frac}}(G^M) = \min\left\{ \sum_{k=1}^p x_k : x \in \mathbb{R}_+^k, \sum_{k:\{i,j\}\subseteq V_k} x_k \ge 1 \text{ for } \{i, j\} \in E^M \right\} \le \text{bc}(G^M). \tag{71}$$

Yet another well-known combinatorial interpretation of bicliques is as follows. Define the rectangular graph $\text{RG}(M)$, with vertex set E^M and where two distinct pairs $\{i, j\}, \{k, \ell\} \in E^M$ form an edge of $\text{RG}(M)$ if $M_{i\ell}M_{kj} = 0$. In other words, $\{i, j\}, \{k, \ell\} \in E^M$ do not form an edge in $\text{RG}(M)$ precisely if $(\{i, k\}, \{j, \ell\})$ corresponds to a biclique in G^M. Then, the parameter $\text{bc}(G^M)$ coincides with the coloring number of $\text{RG}(M)$ and $\text{bc}_{\text{frac}}(G^M)$ coincides with $\chi_f(\text{RG}(M))$, the fractional coloring number of $\text{RG}(M)$. So,

$$\text{rank}_+(M) \ge \text{bc}(G^M) = \chi(\text{RG}(M)).$$

The following relationships are shown in [29]:

$$\tau_+(M) \ge \chi_f(\text{RG}(M)) = \text{bc}_{\text{frac}}(G^M), \quad \tau_+^{\text{sos}}(M) \ge \overline{\vartheta}(\text{RG}(M)),$$

where $\overline{\vartheta}(\text{RG}(M))$ is the theta number of the complement of $\text{RG}(M)$. As we now observe, the ideal-sparse parameter $\xi_1^{+,\text{isp}}(M)$ is at least as good as $\text{bc}_{\text{frac}}(G^M)$, which is the analog of Lemma 13.

Lemma 15 *For $M \in \mathbb{R}_+^{m \times n}$ we have $\xi_1^{+,\text{isp}}(M) \ge \text{bc}_{\text{frac}}(G^M)$.*

Proof Let $(L_1, \ldots, L_p)$ be an optimal solution for $\xi_1^{+,\text{isp}}(M)$. Then, $L_k(M_{i,j-m} - x_i x_j) \ge 0$ for each $k \in [p]$ and $\{i, j\} \in E^M$ such that $\{i, j\} \subseteq V_k$. As $\sum_{k:\{i,j\}\subseteq V_k} L_k(x_i x_j) = M_{i,j-m}$, this implies $\sum_{k:\{i,j\}\subseteq V_k} L_k(1) \ge 1$ for each $\{i, j\} \in E^M$. Hence, the vector $x = (L_k(1))_{k=1}^p$ provides a feasible solution to program (71), which implies $\sum_{k=1}^p L_k(1) \ge \text{bc}_{\text{frac}}(G^M)$. $\qquad\square$

As for the cp-rank, we now give a class of matrices showing a large separation between the ideal-sparse and dense bounds of level $t = 1$.

Example 16 Consider the identity matrix $M = I_n \in \mathcal{S}^n$. Clearly, we have $\text{rank}_{cp}(I_n) = \text{rank}(I_n) = n$. As the support graph G^M is the disjoint union of n edges, its fractional edge biclique-cover number is equal to n and thus, in view of Lemma 15, we have $\xi_1^{+,\text{isp}}(I_n) = n = \text{rank}_+(I_n)$. We now show that for the dense bound, we have $\xi_1^+(I_n) < 8$ for any $n \geq 4$. For this recall that $\xi_1^+(I_n)$ is given by

$$\xi_1^+(I_n) = \min\{L(1) : L \in \mathbb{R}[x]_2^*, \ L(x_i) \geq L(x_i^2) \ (i \in [2n]),$$
$$L(x_i x_{n+j}) = \delta_{i,j} \ (i, j \in [n]), \ L([x]_1[x]_1^T) \succeq 0\}.$$

Consider the linear functional $L \in \mathbb{R}[x]_2^*$ defined by $L(1) = 8\frac{n-2}{n}$, $L(x_i) = L(x_i^2) = 2\frac{n-2}{n}$ for $i \in [2n]$, $L(x_i x_j) = L(x_{n+i} x_{n+j}) = \frac{n-4}{n}$ for $i \neq j \in [n]$, and $L(x_i x_{n+j}) = \delta_{i,j}$ for $i, j \in [n]$. Then one can check that

$$L([x]_1[x]_1^T) = \begin{pmatrix} 8\frac{n-2}{n} & 2\frac{n-2}{n}e^T & 2\frac{n-2}{n}e^T \\ 2\frac{n-2}{n}e & I_n + \frac{n-4}{n}J_n & I_n \\ 2\frac{n-2}{n}e & I_n & I_n + \frac{n-4}{n}J_n \end{pmatrix} \succeq 0.$$

Hence, L is feasible for the program defining $\xi_1^+(I_n)$, which shows $\xi_1^+(I_n) \leq L(1) = 8\frac{n-2}{n} < 8$.

5.3 Numerical results for the nonnegative rank

In this section we test the ideal-sparse and dense hierarchies on two classes of nonnegative matrices. The first class consists of size 4×4 matrices that depend continuously on a single variable. The second class we consider are the Euclidean distance matrices (EDMs).

5.3.1 Matrices related to the nested rectangles problem

The nonnegative matrices we will consider have an interesting link between their nonnegative rank and the geometric nested rectangles problem (see [13]). Bounds for their nonnegative rank were investigated by Fawzi and Parrilo [29] and Gribling et al. [33]. Consider the matrices

$$S(a, b) := \begin{pmatrix} 1-a & 1+a & 1-b & 1+b \\ 1+a & 1-a & 1-b & 1+b \\ 1+a & 1-a & 1+b & 1-b \\ 1-a & 1+a & 1+b & 1-b \end{pmatrix} \quad \text{for } a, b \in [0, 1].$$

If $a, b < 1$, then $S(a, b)$ is fully dense and no improvement can be expected from our new bounds. Thus, we consider the case $b = 1$ and $a \in [0, 1]$. We have computed the
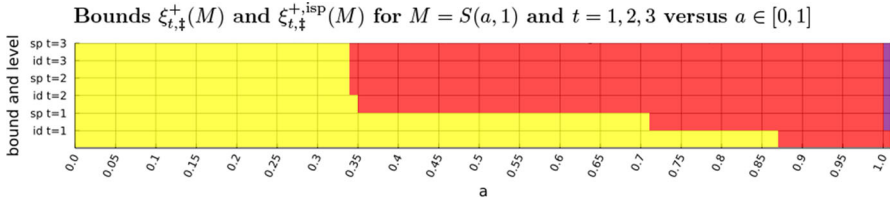
Fig. 3 This figure shows $\xi_{t,\ddagger}^{+}(S(a,1))$ and $\xi_{t,\dagger}^{+,\text{isp}}(S(a,1))$ computed at levels $t = 1, 2, 3$ with a ranging from 0 to 1 in increments of 0.01. The colour indicates a lower bound on the obtained numerical value: yellow, red and purple show the bound is at least 2, 3, and 4, respectively. So a red square at $a = 0.35$ and "sp t=2" means $\xi_{2,\dagger}^{+,\text{isp}}(M) \geq 3$

bounds $\xi_{t,\ddagger}^{+}(M)$ and $\xi_{t,\ddagger}^{+,\text{isp}}(M)$ at level $t = 1, 2, 3$ for $M = S(a, 1)$ with a ranging from 0 to 1 in increments of 0.01. The results are displayed in Fig. 3 below. We can make the following two observations about Fig. 3. First, the ideal-sparse hierarchy is much stronger at level $t = 1$, but at level $t = 2$ the dense and ideal-sparse hierarchies give comparable bounds. Second, for $a = 1$, all bounds, except the dense bound of level 1, are equal to $4 = \text{rank}_+(S(1,1))$ (as is expected for the ideal-sparse hierarchy in view of Lemma 15).

5.3.2 Euclidean distance matrices

The second class of examples we consider are the Euclidean distance matrices $M_n = ((i - i)^2)_{i,j=1}^{n} \in \mathbb{R}_+^{n \times n}$, known to have a large separation between their rank and their nonnegative rank. Indeed, $\text{rank}(M_n) = 3$ [8], and their bipartite support graph G^{M_n} is $K_{n,n}$ with a deleted perfect matching (known as a *crown graph*), whose edge biclique-cover number satisfies $\text{bc}(G^{M_n}) = \Theta(\log n)$ [22]. So we have $\text{rank}(M_n) = 3$ and $\text{rank}_+(M) \geq \text{bc}(G^{M_n}) = \Theta(\log n)$. In addition, it is known that $\text{rank}_+(M_n) \leq 2 + \lceil \frac{n}{2} \rceil$, see [32, Theorem 9]. The numerical results are shown in Table 5. In these examples, the ideal-sparse bound of level $t = 2$ is more difficult to compute, since the support graph G^M has 2^{n-1} maximal bicliques, each with n vertices. For this reason we could compute $\xi_{2,\dagger}^{+,\text{isp}}$ only until $n = 7$ before running out of memory. So this example illustrates the limitations of the ideal-sparsity approach, when the number of maximal cliques is too large. Note that this difficulty – large number of maximal bicliques – remains even if we would replace the support graph G^{M_n} by a supergraph $\tilde{G}$, obtained by adding to M^{G_n} (say) s edges from the missing perfect matching. Indeed, such $\tilde{G}$ still has 2^{n-s-1} maximal bicliques, each with $n + s$ vertices.

6 Concluding remarks

In this paper we have introduced a new sparsity approach for GMP, which arises when in the formulation of GMP one has explicit ideal-type constraints that require the support of the measure to be contained in the variety of an ideal generated by monomials $x_i x_j$ corresponding to (the non-edges of) a graph G. We compared it to the more classic correlative sparsity structure that requires a chordal structure on

Table 5 Bounds for the matrices $M = ((i-j)^2)_{i,j=1}^n$

n	bc	$\xi_{1,\dagger}^+$	$\xi_{2,\dagger}^+$	$\xi_{1,\dagger}^{+,\text{isp}}$	$\xi_{2,\dagger}^{+,\text{isp}}$	$2 + \lceil \frac{n}{2} \rceil$
4	4	2	3.46	3	3.63	4
5	4	2	3.73	3.35	4.19	5
6	4	2	3.96	3.41	4.53	5
7	5	2	4.17	3.55	4.85	6
8	5	2	4.35	3.59	–	6
9	5	2	4.51	3.66	–	7

the graph G, while our new ideal-sparse hierarchy does not need it. We explored its application to the problem of bounding the nonnnegative rank and the cp-rank of a matrix and illustrate the new approach on some classes of examples. There are several natural extensions and further research directions that are left open by this work. We now sketch some of them.

How to deal with many cliques. In the new ideal-sparse approach, instead of a single measure on the full space $\mathbb{R}^n$, one has several measures on smaller spaces indexed by the maximal cliques of the graph G. At any given level $t \geq 1$, the corresponding ideal-sparse bounds are at least as good as their dense analogs and, depending on the number of maximal cliques, their computation can be much faster. The computation of the ideal-sparse parameters indeed involves several (based on the maximal cliques) semidefinite matrices of smaller sizes. The first research direction is to investigate the trade-off between having many cliques (in the ideal-sparse setting) and large matrix constraints (in the dense setting). As seen in Sect. 5.3.2 the sparse hierarchy behaves particularly bad on examples where the underlying graph has exponentially many cliques. We suggest a possible solution in Remark 9, where we consider merging some of the cliques by considering a (possibly chordal) extension of the support graph G. Clique merging has been explored before in the context of power flow networks, see [59] and [31]. These methods exploit correlative sparsity and thus require the underlying support graph to be chordal. Finding the minimal chordal extension of a graph is NP-complete [4], but heuristics exist for certain cases (see, e.g., [10]). Supposing one has chosen a method for finding chordal extensions, it is still unclear which among the possible chordal extensions will result in better SDPs. One can try to merge small cliques based on how much it would reduce the estimated computational burden. These estimates can be based, for example, on the number of constraints, see [52], or on the cost of an interior-point method iteration, see [60]. As it stands, we know of no systematic way to find a "computationally optimal" trade-off between the dense and ideal-sparse hierarchies.

Application to other matrix factorization ranks. We have explored the application to nonnegative and completely positive matrix factorization ranks. We have not considered their non-commutative analogs for the positive semidefinite (psd) rank and the completely positive semidefinite (cpsd) rank, where, respectively, given $M \in \mathbb{R}_+^{m \times n}$ one wants psd matrices $X_i, Y_j \in \mathcal{S}^r$ such that $M = (\langle X_i, Y_j \rangle)_{i \in [m], j \in [n]}$, and given

$A \in \mathcal{S}^n$ one wants psd matrices $X_i \in \mathcal{S}_+^r$ such that $A = (\langle X_i, X_j \rangle)_{i,j \in [n]}$, with r smallest possible. One recovers the nonnegative rank and the cp-rank when restricting the factors X_i, Y_j to be diagonal matrices. We refer the reader to [33], where a common polynomial optimization framework is offered to treat all these four matrix factorization ranks. In the noncommutative setting of the psd- and cpsd-ranks, zero entries of M (or A) also imply ideal-type constraints of the form $X_i Y_j = 0$ (or $X_i X_j = 0$). Thus the techniques in the present paper may extend to this general setting. We leave this extension to future work.

More general ideal-sparsity and applications. We have considered an ideal-sparsity structure, where the ideal in (3) is generated by quadratic monomials. Beside their use for bounding matrix factorization ranks, constraints of the form $x_i x_j = 0$ naturally arise in a number of other applications. First we note that up to a change of variables, one can consider more general constraints of the form $(a^\top x + b)(c^\top x + d) = 0$. This type of constraint is commonly referred to as a *complementarity constraint*, where either the term $(a^\top x + b)$ or the term $(c^\top x + d)$ is required to be zero. We mention two areas where such complementarity constraints naturally arise: analysis of neural networks and optimality conditions in optimization.

Complementarity constraints arise naturally when modeling neural networks with the rectified linear activation functions (ReLU). The semialgebraic representation of the graph of the ReLU function involves a constraint of the form $y(y - x) = 0$, which is exactly a complementarity constraint. The fact that the graph of the ReLU function admits a semialgebraic representation has been exploited computationally using the moment-sum-of-squares framework, for analyzing the Lipschitz constant of the neural network as well as stability and performance properties of dynamical systems controlled by the ReLU neural networks, see, e.g., [16, 17, 41]. Ideal sparsity is therefore a natural candidate to render these methods more computationally efficient and would deserve further study.

Complementarity systems arise also in optimization within the Karush-Kuhn-Tucker (KKT) conditions. The complementarity slackness of the KKT condition reads $\lambda_i f_i(x) = 0$, where λ_i is the Lagrange multiplier associated to the ith constraint $f_i(x) \leq 0$. If f_i is affine, this is in the form of ideal constraints. The fact that the KKT conditions form a basic semialgebraic set when the optimization problem has polynomial data was exploited in [42] to analyze dynamical systems controlled by optimization algorithms, albeit without exploiting the ideal-sparsity. More generally, the ideal-sparsity could be used to analyze the *linear complementarity problems* (LCP) that have applications in, e.g., economics, engineering or game theory; see [18] for an extensive treatment of the subject.

Finally, instead of considering an ideal generated by quadratic monomials, one may consider an ideal generated by a set of monomials $x^S = \prod_{i \in S} x_i$ $(S \in \mathcal{S})$, where $\mathcal{S}$ is a given collection of subsets of $V = [n]$. The treatment extends naturally to this more general setting, where in the definition (2) of the set K, we replace the constraints $x_i x_j = 0$ $(\{i, j\} \in \overline{E})$ by $\prod_{i \in S} x_i = 0$ $(S \in \mathcal{S})$. Indeed, let $V_1, \ldots, V_p$ denote the maximal subsets of V that do not contain any set $S \in \mathcal{S}$. Then, for the dense formulation (1) of GMP, one can again show an equivalent sparse reformulation as in (11), which involves p measures supported on the subspaces $\mathbb{R}^{|V_1|}, \ldots, \mathbb{R}^{|V_p|}$ instead

of a single measure on $\mathbb{R}^{|V|}$. We leave it for further research to explore applications of this more general ideal-sparsity setting and possible further extensions to other types of varieties.

References

1. Abdalmoaty, M.R., Henrion, D., Rodrigues. L.: Measures and LMIs for optimal control of piecewise-affine systems. In: 2013 European Control Conference (ECC), pp. 3173–3178. IEEE (2013)
2. Agler, J., Helton, W., McCullough, S.A., Rodman, L.: Positive semidefinite matrices with a given sparsity pattern. Linear Algebra Appl. **107**, 101–149 (1988)
3. Andersen, E.D., Andersen, K.D.: The Mosek interior point optimizer for linear programming: an implementation of the homogeneous algorithm. In: High Performance Optimization, vol. 33, pp. 197–232. Springer (2000)
4. Arnborg, S., Corneil, D.G., Proskurowski, A.: Complexity of finding embeddings in a k-tree. SIAM J. Algebraic Discrete Methods **8**(2), 277–284 (1987)
5. Averkov, G.: Optimal size of linear matrix inequalities in semidefinite approaches to polynomial optimization. SIAM J. Appl. Algebra Geom. **3**(1), 128–151 (2019)
6. Barioli, F.: Completely positive matrices with a book-graph. Linear Algebra Appl. **277**(1), 11–31 (1998)
7. Barioli, F., Berman, A.: The maximal cp-rank of rank k completely positive matrices. Linear Algebra Appl. **363**, 17–33 (2003)
8. Beasley, L.B., Laffey, T.J.: Real rank versus nonnegative rank. Linear Algebra Appl. **431**, 2330–2335 (2009)
9. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: a fresh approach to numerical computing. SIAM Rev. **59**, 65–98 (2017)
10. Bodlaender, H.L., Koster, A.M.: Treewidth computations I: upper bounds. Inf. Comput. **208**(3), 259–275 (2010)
11. Bomze, I.M., Schachinger, W., Ullrich, R.: From seven to eleven: completely positive matrices with high cp-rank. Linear Algebra Appl. **459**, 208–221 (2014)
12. Bomze, I.M., Schachinger, W., Ullrich, R.: New lower bounds and asymptotics for the cp-rank. SIAM J. Matrix Anal. Appl. **36**, 20–37 (2015)
13. Braun, G., Fiorini, S., Pokutta, S., Steurer, D.: Approximation limits of linear programs (beyond hierarchies). Math. Oper. Res. **40**(3), 756–772 (2015)
14. Bron, C., Kerbosch, J.: Algorithm 457: finding all cliques of an undirected graph. Commun. ACM **16**, 575–577 (1973)
15. Chandrasekaran, V., Shah, P.: Relative entropy relaxations for signomial optimization. SIAM J. Optim. **26**(2), 1147–1173 (2016)
16. Chen, T., Lasserre, J.B., Magron, V., Pauwels. E.: Semialgebraic optimization for bounding Lipschitz constants of ReLU networks. In: 34th Conference on Neural Information Processing Systems (NeurIPS 2020) (2020)
17. Chen, T., Lasserre, J.-B., Magron, V., Pauwels, E.: Semialgebraic representation of monotone deep Equilibrium models and applications to certification. In: 35th Conference on Neural Information Processing Systems (NeurIPS 2021) (2021)

18. Cottle, R.W., Pang, J.-S., Stone, R.E.: The Linear Complementarity Problem. SIAM (2009)
19. Curto, R.E., Fialkow, L.A.: Solution of the truncated complex moment problem for flat data. Mem. Am. Math. Soc., **119**(568) (1996)
20. Curto, R.E., Fialkow, L.A.: The truncated complex moment problem. Trans. Am. Math. Soc. **352**, 2825–2855 (2000)
21. de Klerk, E., Laurent, M.: A survey of semidefinite programming approaches to the generalized problem of moments and their error analysis. In: World Women in Mathematics 2018, Araujo C., Benkart G., Praeger C., Tanbay B. (eds). Association for Women in Mathematics Series, Springer, Cham, vol. 20, pp. 17–56 (2019)
22. de Caen, D., Gregory, D.A., Pullman, N.J.: The boolean rank of zero-one matrices. In: Proceedings of the Third Caribbean Conference on Combinatorics and Computing, Barbados, pp. 169–173 (1981)
23. de Klerk, E., Vallentin, F.: On the Turing model complexity of interior point methods for semidefinite programming. SIAM J. Optim. **26**(3), 1944–1961 (2016)
24. Dickinson, P.J.C., Gijben, L.: On the computational complexity of membership problems for the completely positive cone and its dual. Comput. Optim. Appl. **57**, 403–415 (2014)
25. Diestel, R.: Graph Theory. Springer, Berlin, Heidelberg (2017)
26. Drew, J.H., Johnson, C.R., Loewy, R.: Completely positive matrices associated with M-matrices. Linear Multilinear Algebra **37**(4), 303–310 (1994)
27. Dunning, I., Huchette, J., Lubin, M.: JuMP: a modeling language for mathematical optimization. SIAM Rev. **59**(2), 295–320 (2017)
28. Erdös, P., Goodman, A.W., Pósa, L.: The representation of a graph by set intersections. Can. J. Math. **18**, 106–112 (1966)
29. Fawzi, H., Parrilo, P.A.: Self-scaled bounds for atomic cone ranks: applications to nonnegative rank and cp-rank. Math. Program. **158**, 417–465 (2016)
30. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman & Company, Publishers, San Francisco (1978)
31. Garstka, M., Cannon, M., Goulart, P.: A clique graph based merging strategy for decomposable SDPs. IFAC-PapersOnLine **53**(2), 7355–7361 (2020)
32. Gillis, N., Glineur, F.: On the geometric interpretation of the nonnegative rank. Linear Algebra Appl. **437**, 2685–2712 (2012)
33. 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)
34. Gribling, S., Laurent, M., Steenkamp, A.: Bounding the separable rank via polynomial optimization. Linear Algebra Appl. **648**, 1–55 (2022)
35. Grimm, D., Netzer, T., Schweighofer, M.: A note on the representation of positive polynomials with structured sparsity. Arch. Math. **89**, 399–403 (2007)
36. Henrion, D., Korda, M., Lasserre, J.B.: The Moment-SOS Hierarchy, volume 4. World Scientific (2020)
37. Henrion, D., Lasserre, J.-B.: Detecting global optimality and extracting solutions in GloptiPoly. In: Positive Polynomials in Control, D. Henrion and A. Garulli (eds.), Lecture Notes on Control and Information Sciences, 312:293–310, Springer, Berlin (2005)
38. Iliman, S., De Wolff, T.: Amoebas, nonnegative polynomials and sums of squares supported on circuits. Res. Math. Sci. **3**(1), 1–35 (2016)
39. 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)
40. Klep, I., Magron, V., Povh, J.: Sparse noncommutative polynomial optimization. Math. Program. **193**, 1–41 (2021)
41. Korda, M.: Stability and performance verification of dynamical systems controlled by neural networks: algorithms and complexity. IEEE Control Syst. Lett. **6**, 3265–3270 (2022)
42. Korda, M., Jones, C.N.: Stability and performance verification of optimization-based controllers. Automatica **78**, 34–45 (2017)
43. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**, 796–817 (2001)
44. Lasserre, J.B.: Convergent SDP-relaxations in polynomial optimization with sparsity. SIAM J. Optim. **17**(3), 822–843 (2006)
45. Lasserre, J.B.: A semidefinite programming approach to the generalized problem of moments. Math. Program. **112**, 65–92 (2008)
46. Lasserre, J.B.: Moments. Imperial College Press, Positive Polynomials and Their Applications (2009)

47. Lasserre, J.B.: An Introduction to Polynomial and Semi-Algebraic Optimization, Cambridge University Press (2015)
48. Lasserre, J.B., Emin, Y.: Semidefinite relaxations for Lebesgue and Gaussian measures of unions of basic semialgebraic sets. Math. Oper. Res. **44**(4), 1477–1493 (2019)
49. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. In: Emerging Applications of Algebraic Geometry, vol. 149 of IMA Volumes in Mathematics and its Applications, M. Putinar and S. Sullivant (eds.), Springer, pp. 157–270 (2009)
50. Magron, V., Forets, M., Henrion, D.: Semidefinite approximations of invariant measures for polynomial systems. Discrete Contin. Dyn. Syst. B **24**(12), 6745–6770 (2019)
51. Magron, V., Wang, J.: Sparse Polynomial Optimization: Theory and Practice. World Scientific Press, Series on Optimization and Its Applications (2023)
52. Molzahn, D.K., Holzer, J.T., Lesieutre, B.C., DeMarco, C.L.: Implementation of a large-scale optimal power flow solver based on semidefinite programming. IEEE Trans. Power Syst. **28**(4), 3987–3998 (2013)
53. Nie, J.: The $\mathcal{A}$-truncated K-moment problem. Found. Comput. Math. **14**(6), 1243–1276 (2014)
54. Nie, J.: Symmetric tensor nuclear norms. SIAM J. Appl. Algebra Geom. **1**(1), 599–625 (2017)
55. Reznick, B.: Extremal PSD forms with few terms. Duke Math. J. **45**(2), 363–374 (1978)
56. Riener, C., Theobald, T., Andrén, L.J., Lasserre, J.B.: Exploiting symmetries in SDP-relaxations for polynomial optimization. Math. Oper. Res. **38**(1), 122–141 (2013)
57. Salce, L., Zanardo, P.: Completely positive matrices and positivity of least squares solutions. Linear Algebra Appl. **178**, 201–216 (1993)
58. Shaked-Monderer, N., Bomze, I.M., Jarre, F., Schachinger, W.: On the cp-rank and minimal cp factor-izations of a completely positive matrix. SIAM J. Matrix Anal. Appl. **34**(2), 355–368 (2013)
59. Sliwak, J., Andersen, E.D., Anjos, M.F., Létocart, L., Traversi, E.: A clique merging algorithm to solve semidefinite relaxations of optimal power flow problems. IEEE Trans. Power Syst. **36**(2), 1641–1644 (2021)
60. Sliwak, J., Anjos, M.F., Létocart, L., Traversi, E.: A semidefinite optimization-based branch-and-bound algorithm for several reactive optimal power flow problems. arXiv:2103.13648 (2021)
61. Tang, G., Shah, P.: Guaranteed tensor decomposition: A moment approach. In: Bach, F., Blei, D., (eds). Proceedings of the 32nd International Conference on Machine Learning, volume 37 of Proceedings of Machine Learning Research, PMLR 37:1491–1500 (2015)
62. Vavasis, S.A.: On the complexity of nonnegative matrix factorization. SIAM J. Optim. **20**(3), 1364–1377 (2010)
63. Waki, H., Kim, S., Kojima, M., Muramatsu, M.: Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. SIAM J. Optim. **17**(1), 218–242 (2006)
64. Wang, J., Magron, V.: A second order cone characterization for sums of nonnegative circuits. In: International Symposium on Symbolic and Algebraic Computation (ISSAC '20), July 20–23, 2020, Kalamata, Greece, pp. 450–457. ACM, New York, NY, USA
65. 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)
66. Wang, J., Magron, V., Lasserre, J.B.: TSSOS: a moment-SOS hierarchy that exploits term sparsity. SIAM J. Optim. **31**(1), 30–58 (2021)
67. Wang, J., Magron, V., Lasserre, J.B., Mai, N.H.A.: CS-TSSOS: correlative and term sparsity for large-scale polynomial optimization. ACM Trans. Math. Softw. **48**(4), 1–26 (2022)
68. Xiang, S., Xiang, S.: Notes on completely positive matrices. Linear Algebra Appl. **271**(1), 273–282 (1998)
69. Zheng, Y., Fantuzzi, G.: Sum-of-squares chordal decomposition of polynomial matrix inequalities. Math. Program. **197**, 71–108 (2023)
70. Zheng, Y., Fantuzzi, G., Papachristodoulou, A.: Chordal and factor-width decompositions for scalable semidefinite and polynomial optimization. Annu. Rev. Control. **52**, 243–279 (2021)