# Lower Bounds on Matrix Factorization Ranks via Noncommutative Polynomial Optimization

Sander Gribling[1] · David de Laat[1] · Monique Laurent[1,2]

## Abstract
We use techniques from (tracial noncommutative) polynomial optimization to formulate hierarchies of semidefinite programming lower bounds on matrix factorization ranks. In particular, we consider the nonnegative rank, the positive semidefinite rank, and their symmetric analogs: the completely positive rank and the completely positive semidefinite rank. We study convergence properties of our hierarchies, compare them extensively to known lower bounds, and provide some (numerical) examples.

## 1 Introduction

### 1.1 Matrix Factorization Ranks

A factorization of a matrix $A \in \mathbb{R}^{m \times n}$ over a sequence $\{K^d\}_{d \in \mathbb{N}}$ of cones that are each equipped with an inner product $\langle \cdot, \cdot \rangle$ is a decomposition of the form $A = (\langle X_i, Y_j \rangle)$ with $X_i, Y_j \in K^d$ for all $(i, j) \in [m] \times [n]$, for some integer $d \in \mathbb{N}$. Following [34], the

---

✉ Monique Laurent
   M.Laurent@cwi.nl

1   CWI, Amsterdam, The Netherlands

2   Tilburg University, Tilburg, The Netherlands

smallest integer $d$ for which such a factorization exists is called the *cone factorization rank* of $A$ over $\{K^d\}$.

The cones $K^d$ we use in this paper are the nonnegative orthant $\mathbb{R}^d_+$ with the usual inner product and the cone $S^d_+$ (resp., $H^d_+$) of $d \times d$ real symmetric (resp., Hermitian) positive semidefinite matrices with the trace inner product $\langle X, Y \rangle = \mathrm{Tr}(X^\mathsf{T} Y)$ (resp., $\langle X, Y \rangle = \mathrm{Tr}(X^* Y)$). We obtain the *nonnegative rank*, denoted $\mathrm{rank}_+(A)$, which uses the cones $K^d = \mathbb{R}^d_+$, and the *positive semidefinite rank*, denoted $\mathrm{psd\text{-}rank}_\mathbb{K}(A)$, which uses the cones $K^d = S^d_+$ for $\mathbb{K} = \mathbb{R}$ and $K^d = H^d_+$ for $\mathbb{K} = \mathbb{C}$. Both the nonnegative rank and the positive semidefinite rank are defined whenever $A$ is entrywise nonnegative.

The study of the nonnegative rank is largely motivated by the groundbreaking work of Yannakakis [78], who showed that the linear extension complexity of a polytope $P$ is given by the nonnegative rank of its slack matrix. The *linear extension complexity* of $P$ is the smallest integer $d$ for which $P$ can be obtained as the linear image of an affine section of the nonnegative orthant $\mathbb{R}^d_+$. The *slack matrix* of $P$ is given by the matrix $(b_i - a_i^T v)_{v \in V, i \in I}$, where $P = \mathrm{conv}(V)$ and $P = \{x : a_i^T x \leq b_i \ (i \in I)\}$ are the point and hyperplane representations of $P$. Analogously, the *semidefinite extension complexity* of $P$ is the smallest $d$ such that $P$ is the linear image of an affine section of the cone $S^d_+$ and it is given by the (real) positive semidefinite rank of its slack matrix [34].

The motivation to study the linear and semidefinite extension complexities is that polytopes with small extension complexity admit efficient algorithms for linear optimization. Well-known examples include spanning tree polytopes [54] and permutahedra [32], which have polynomial linear extension complexity, and the stable set polytope of perfect graphs, which has polynomial semidefinite extension complexity [40] (see, e.g., the surveys [18,25]). The above connection to the nonnegative rank and to the positive semidefinite rank of the slack matrix can be used to show that a polytope does not admit a small extended formulation. Recently, this connection was used to show that the linear extension complexities of the traveling salesman, cut, and stable set polytopes are exponential in the number of nodes [29], and this result was extended to their semidefinite extension complexities in [51]. Surprisingly, the linear extension complexity of the matching polytope is also exponential [66], even though linear optimization over this set is polynomial time solvable [23]. It is an open question whether the semidefinite extension complexity of the matching polytope is exponential.

Besides this link to extension complexity, the nonnegative rank also finds applications in probability theory and in communication complexity, and the positive semidefinite rank has applications in quantum information theory and in quantum communication complexity (see, e.g., [24,29,42,55]).

For square symmetric matrices ($m = n$), we are also interested in *symmetric* analogs of the above matrix factorization ranks, where we require the same factors for the rows and columns (i.e., $X_i = Y_i$ for all $i \in [n]$). The symmetric analog of the nonnegative rank is the *completely positive rank*, denoted $\mathrm{cp\text{-}rank}(A)$, which uses the cones $K^d = \mathbb{R}^d_+$, and the symmetric analog of the positive semidefinite rank is the *completely positive semidefinite rank*, denoted $\mathrm{cpsd\text{-}rank}_\mathbb{K}(A)$, which uses the cones

$K^d = S_+^d$ if $\mathbb{K} = \mathbb{R}$ and $K^d = H_+^d$ if $\mathbb{K} = \mathbb{C}$. These symmetric factorization ranks are not always well defined since not every symmetric nonnegative matrix admits a symmetric factorization by nonnegative vectors or positive semidefinite matrices. The symmetric matrices for which these parameters are well defined form convex cones known as the *completely positive cone*, denoted $\mathrm{CP}^n$, and the *completely positive semidefinite cone*, denoted $\mathrm{CS}_+^n$. We have the inclusions $\mathrm{CP}^n \subseteq \mathrm{CS}_+^n \subseteq S_+^n$, which are known to be strict for $n \geq 5$. For details on these cones see [6,17,50] and references therein.

Motivation for the cones $\mathrm{CP}^n$ and $\mathrm{CS}_+^n$ comes in particular from their use to model classical and quantum information optimization problems. For instance, graph parameters such as the stability number and the chromatic number can be written as linear optimization problems over the completely positive cone [45], and the same holds, more generally, for quadratic problems with mixed binary variables [13]. The cp-rank is widely studied in the linear algebra community; see, e.g., [6,10,68,69].

The completely positive semidefinite cone was first studied in [50] to describe quantum analogs of the stability number and of the chromatic number of a graph. This was later extended to general graph homomorphisms in [72] and to graph isomorphism in [2]. In addition, as shown in [53,72], there is a close connection between the completely positive semidefinite cone and the set of quantum correlations. This also gives a relation between the completely positive semidefinite rank and the minimal entanglement dimension necessary to realize a quantum correlation. This connection has been used in [38,62,63] to construct matrices whose completely positive semidefinite rank is exponentially large in the matrix size. For the special case of synchronous quantum correlations, the minimum entanglement dimension is directly given by the completely positive semidefinite rank of a certain matrix (see [37]).

The following inequalities hold for the nonnegative rank and the positive semidefinite rank: We have

$$\text{psd-rank}_{\mathbb{C}}(A) \leq \text{psd-rank}_{\mathbb{R}}(A) \leq \text{rank}_+(A) \leq \min\{m, n\}$$

for any $m \times n$ nonnegative matrix $A$ and $\text{cp-rank}(A) \leq \binom{n+1}{2}$ for any $n \times n$ completely positive matrix $A$. However, the situation for the cpsd-rank is very different. Exploiting the connection between the completely positive semidefinite cone and quantum correlations it follows from results in [73] that the cone $\mathrm{CS}_+^n$ is not closed for $n \geq 1942$. The results in [22] show that this already holds for $n \geq 10$. As a consequence, there does not exist an upper bound on the cpsd-rank as a function of the matrix size. For small matrix sizes, very little is known. It is an open problem whether $\mathrm{CS}_+^5$ is closed, and we do not even know how to construct a $5 \times 5$ matrix whose cpsd-rank exceeds 5.

The rank$_+$, cp-rank, and psd-rank are known to be computable; this follows using results from [65] since upper bounds exist on these factorization ranks that depend only on the matrix size, see [5] for a proof for the case of the cp-rank. But computing the nonnegative rank is NP-hard [76]. In fact, determining the rank$_+$ and psd-rank of a matrix are both equivalent to the existential theory of the reals [70,71]. For the cp-rank and the cpsd-rank, no such results are known, but there is no reason to assume they are any easier. In fact, it is not even clear whether the cpsd-rank is computable in general.

To obtain upper bounds on the factorization rank of a given matrix, one can employ heuristics that try to construct small factorizations. Many such heuristics exist for the nonnegative rank (see the overview [30] and references therein), factorization algorithms exist for completely positive matrices (see the recent paper [39], also [20] for structured completely positive matrices), and algorithms to compute positive semidefinite factorizations are presented in the recent work [75]. In this paper, we want to compute *lower bounds* on matrix factorization ranks, which we achieve by employing a relaxation approach based on (noncommutative) polynomial optimization.

## 1.2 Contributions and Connections to Existing Bounds

In this work, we provide a unified approach to obtain lower bounds on the four matrix factorization ranks mentioned above, based on tools from (noncommutative) polynomial optimization.

We sketch the main ideas of our approach in Sect. 1.4 below, after having introduced some necessary notation and preliminaries about (noncommutative) polynomials in Sect. 1.3. We then indicate in Sect. 1.5 how our approach relates to the more classical use of polynomial optimization dealing with the minimization of polynomials over basic closed semialgebraic sets. The main body of the paper consists of four sections each dealing with one of the four matrix factorization ranks. We start with presenting our approach for the completely positive semidefinite rank and then explain how to adapt this to the other ranks.

For our results, we need several technical tools about linear forms on spaces of polynomials, both in the commutative and noncommutative setting. To ease the readability of the paper, we group these technical tools in Appendix A. Moreover, we provide full proofs, so that our paper is self-contained. In addition, some of the proofs might differ from the customary ones in the literature since our treatment in this paper is consistently on the 'moment' side rather than using real algebraic results about sums of squares.

In Sect. 2, we introduce our approach for the completely positive semidefinite rank. We start by defining a hierarchy of lower bounds

$$\xi_1^{\text{cpsd}}(A) \le \xi_2^{\text{cpsd}}(A) \le \ldots \le \xi_t^{\text{cpsd}}(A) \le \ldots \le \text{cpsd-rank}_{\mathbb{C}}(A),$$

where $\xi_t^{\text{cpsd}}(A)$, for $t \in \mathbb{N}$, is given as the optimal value of a semidefinite program whose size increases with $t$. Not much is known about lower bounds for the cpsd-rank in the literature. The inequality $\sqrt{\text{rank}(A)} \le \text{cpsd-rank}_{\mathbb{C}}(A)$ is known, which follows by viewing a Hermitian $d \times d$ matrix as a $d^2$-dimensional real vector, and an analytic lower bound is given in [62]. We show that the new parameter $\xi_1^{\text{cpsd}}(A)$ is at least as good as this analytic lower bound and we give a small example where a strengthening of $\xi_2^{\text{cpsd}}(A)$ is strictly better then both above-mentioned generic lower bounds. Currently, we lack evidence that the lower bounds $\xi_t^{\text{cpsd}}(A)$ can be larger than, for example, the matrix size, but this could be because small matrices with large

cpsd-rank are hard to construct or might even not exist. We also introduce several ideas leading to strengthenings of the basic bounds $\xi_t^{\mathrm{cpsd}}(A)$.

We then adapt these ideas to the other three matrix factorization ranks discussed above, where for each of them we obtain analogous hierarchies of bounds.

For the nonnegative rank and the completely positive rank, much more is known about lower bounds. The best-known generic lower bounds are due to Fawzi and Parrilo [26,27]. In [27], the parameters $\tau_+(A)$ and $\tau_{\mathrm{cp}}(A)$ are defined, which, respectively, lower bound the nonnegative rank and the cp-rank, along with their computable semidefinite programming relaxations $\tau_+^{\mathrm{sos}}(A)$ and $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$. In [27] it is also shown that $\tau_+(A)$ is at least as good as certain norm-based lower bounds. In particular, $\tau_+(\cdot)$ is at least as good as the $\ell_\infty$ norm-based lower bound, which was used by Rothvoß [66] to show that the matching polytope has exponential linear extension complexity. In [26] it is shown that for the Frobenius norm, the square of the norm-based bound is still a lower bound on the nonnegative rank, but it is not known how this lower bound compares to $\tau_+(\cdot)$.

Fawzi and Parrilo [27] use the atomicity of the nonnegative and completely positive ranks to derive the parameters $\tau_+(A)$ and $\tau_{\mathrm{cp}}(A)$; i.e., they use the fact that the nonnegative rank (cp-rank) of $A$ is equal to the smallest $d$ for which $A$ can be written as the sum of $d$ nonnegative (positive semidefinite) rank one matrices. As the psd-rank and cpsd-rank are not known to admit atomic formulations, the techniques from [27] do not extend directly to these factorization ranks. However, our approach via polynomial optimization captures these factorization ranks as well.

In Sects. 3 and 4, we construct semidefinite programming hierarchies of lower bounds $\xi_t^{\mathrm{cp}}(A)$ and $\xi_t^+(A)$ on cp-rank$(A)$ and rank$_+(A)$. We show that the bounds $\xi_t^+(A)$ converge to $\tau_+(A)$ as $t \to \infty$. The basic hierarchy $\{\xi_t^{\mathrm{cp}}(A)\}$ for the cp-rank does not converge to $\tau_{\mathrm{cp}}(A)$ in general, but we provide two types of additional constraints that can be added to the program defining $\xi_t^{\mathrm{cp}}(A)$ to ensure convergence to $\tau_{\mathrm{cp}}(A)$. First, we show how a generalization of the tensor constraints that are used in the definition of the parameter $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$ can be used for this, and we also give a more efficient (using smaller matrix blocks) description of these constraints. This strengthening of $\xi_2^{\mathrm{cp}}(A)$ is then at least as strong as $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$, but requires matrix variables of roughly half the size. Alternatively, we show that for every $\varepsilon > 0$ there is a finite number of additional linear constraints that can be added to the basic hierarchy $\{\xi_t^{\mathrm{cp}}(A)\}$ so that the limit of the sequence of these new lower bounds $\xi_t^+(A)$ is at least $\tau_{\mathrm{cp}}(A) - \varepsilon$. We give numerical results on small matrices studied in the literature, which show that $\xi_3^+(A)$ can improve over $\tau_+^{\mathrm{sos}}(A)$.

Finally, in Sect. 5, we derive a hierarchy $\{\xi_t^{\mathrm{psd}}(A)\}$ of lower bounds on the psd-rank. We compare the new bounds $\xi_t^{\mathrm{psd}}(A)$ to a bound from [52], and we provide some numerical examples illustrating their performance.

We provide two implementations of all the lower bounds introduced in this paper, at the arXiv submission of this paper. One implementation uses Matlab and the CVX package [36], and the other one uses Julia [8]. The implementations support various semidefinite programming solvers, for our numerical examples we used Mosek [56].

### 1.3 Preliminaries

In order to explain our basic approach in the next section, we first need to introduce some notation. We denote the set of all words in the symbols $x_1, \ldots, x_n$ by $\langle \mathbf{x} \rangle = \langle x_1, \ldots, x_n \rangle$, where the empty word is denoted by 1. This is a semigroup with involution, where the binary operation is concatenation, and the involution of a word $w \in \langle \mathbf{x} \rangle$ is the word $w^*$ obtained by reversing the order of the symbols in $w$. The $*$-algebra of all real linear combinations of these words is denoted by $\mathbb{R}\langle \mathbf{x} \rangle$, and its elements are called *noncommutative polynomials*. The involution extends to $\mathbb{R}\langle \mathbf{x} \rangle$ by linearity. A polynomial $p \in \mathbb{R}\langle \mathbf{x} \rangle$ is called *symmetric* if $p^* = p$ and $\mathrm{Sym}\,\mathbb{R}\langle \mathbf{x} \rangle$ denotes the set of symmetric polynomials. The degree of a word $w \in \langle \mathbf{x} \rangle$ is the number of symbols composing it, denoted as $|w|$ or $\deg(w)$, and the degree of a polynomial $p = \sum_w p_w w \in \mathbb{R}\langle \mathbf{x} \rangle$ is the maximum degree of a word $w$ with $p_w \neq 0$. Given $t \in \mathbb{N} \cup \{\infty\}$, we let $\langle \mathbf{x} \rangle_t$ be the set of words $w$ of degree $|w| \leq t$, so that $\langle \mathbf{x} \rangle_\infty = \langle \mathbf{x} \rangle$, and $\mathbb{R}\langle \mathbf{x} \rangle_t$ is the real vector space of noncommutative polynomials $p$ of degree $\deg(p) \leq t$. Given $t \in \mathbb{N}$, we let $\langle \mathbf{x} \rangle_{=t}$ be the set of words of degree exactly equal to $t$.

For a set $S \subseteq \mathrm{Sym}\,\mathbb{R}\langle \mathbf{x} \rangle$ and $t \in \mathbb{N} \cup \{\infty\}$, the *truncated quadratic module* at degree $2t$ associated to $S$ is defined as the cone generated by all polynomials $p^* g p \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}$ with $g \in S \cup \{1\}$:

$$\mathscr{M}_{2t}(S) = \mathrm{cone}\left\{ p^* g p : p \in \mathbb{R}\langle \mathbf{x} \rangle, \ g \in S \cup \{1\}, \ \deg(p^* g p) \leq 2t \right\}. \tag{1}$$

Likewise, for a set $T \subseteq \mathbb{R}\langle \mathbf{x} \rangle$, we can define the *truncated ideal* at degree $2t$, denoted by $\mathscr{I}_{2t}(T)$, as the vector space spanned by all polynomials $ph \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}$ with $h \in T$:

$$\mathscr{I}_{2t}(T) = \mathrm{span}\left\{ ph : p \in \mathbb{R}\langle \mathbf{x} \rangle, \ h \in T, \ \deg(ph) \leq 2t \right\}. \tag{2}$$

We say that $\mathscr{M}(S) + \mathscr{I}(T)$ is *Archimedean* when there exists a scalar $R > 0$ such that

$$R - \sum_{i=1}^{n} x_i^2 \in \mathscr{M}(S) + \mathscr{I}(T). \tag{3}$$

Throughout, we are interested in the space $\mathbb{R}\langle \mathbf{x} \rangle_t^*$ of real-valued linear functionals on $\mathbb{R}\langle \mathbf{x} \rangle_t$. We list some basic definitions: A linear functional $L \in \mathbb{R}\langle \mathbf{x} \rangle_t^*$ is *symmetric* if $L(w) = L(w^*)$ for all $w \in \langle \mathbf{x} \rangle_t$ and *tracial* if $L(ww') = L(w'w)$ for all $w, w' \in \langle \mathbf{x} \rangle_t$. A linear functional $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ is said to be *positive* if $L(p^* p) \geq 0$ for all $p \in \mathbb{R}\langle \mathbf{x} \rangle_t$. Many properties of a linear functional $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ can be expressed as properties of its associated moment matrix (also known as its *Hankel matrix*). For $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ we define its associated *moment matrix*, which has rows and columns indexed by words in $\langle \mathbf{x} \rangle_t$, by

$$M_t(L)_{w,w'} = L(w^* w') \quad \text{for} \quad w, w' \in \langle \mathbf{x} \rangle_t,$$

and as usual we set $M(L) = M_\infty(L)$. It then follows that $L$ is symmetric if and only if $M_t(L)$ is symmetric, and $L$ is positive if and only if $M_t(L)$ is positive semidefinite.

In fact, one can even express nonnegativity of a linear form $L \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}^*$ on $\mathscr{M}_{2t}(S)$ in terms of certain associated positive semidefinite moment matrices. For this, given a polynomial $g \in \mathbb{R}\langle\mathbf{x}\rangle$, define the linear form $gL \in \mathbb{R}\langle\mathbf{x}\rangle_{2t-\deg(g)}^*$ by $(gL)(p) = L(gp)$. Then, we have

$$L(p^*gp) \geq 0 \text{ for all } p \in \mathbb{R}\langle\mathbf{x}\rangle_{t-d_g} \iff M_{t-d_g}(gL) \succeq 0, \quad (d_g = \lceil \deg(g)/2 \rceil),$$

and thus $L \geq 0$ on $\mathscr{M}_{2t}(S)$ if and only if $M_{t-d_g}(gL) \succeq 0$ for all $g \in S \cup \{1\}$. Also, the condition $L = 0$ on $\mathscr{I}_{2t}(T)$ corresponds to linear equalities on the entries of $M_t(L)$.

The moment matrix also allows us to define a property called *flatness*. For $t \in \mathbb{N}$, a linear functional $L \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}^*$ is called $\delta$-*flat* if the rank of $M_t(L)$ is equal to that of its principal submatrix indexed by the words in $\langle\mathbf{x}\rangle_{t-\delta}$, that is,

$$\text{rank}(M_t(L)) = \text{rank}(M_{t-\delta}(L)). \tag{4}$$

We call $L$ *flat* if it is $\delta$-flat for some $\delta \geq 1$. When $t = \infty$, $L$ is said to be *flat* when $\text{rank}(M(L)) < \infty$, which is equivalent to $\text{rank}(M(L)) = \text{rank}(M_s(L))$ for some $s \in \mathbb{N}$.

A key example of a flat symmetric tracial positive linear functional on $\mathbb{R}\langle\mathbf{x}\rangle$ is given by the *trace evaluation* at a given matrix tuple $\mathbf{X} = (X_1, \ldots, X_n) \in (\mathbf{H}^d)^n$:

$$p \mapsto \text{Tr}(p(\mathbf{X})).$$

Here, $p(\mathbf{X})$ denotes the matrix obtained by substituting $x_i$ by $X_i$ in $p$, and throughout $\text{Tr}(\cdot)$ denotes the usual matrix trace, which satisfies $\text{Tr}(I) = d$ where $I$ is the identity matrix in $\mathbf{H}^d$. We mention in passing that we use $\text{tr}(\cdot)$ to denote the *normalized matrix trace*, which satisfies $\text{tr}(I) = 1$ for $I \in \mathbf{H}^d$. Throughout, we use $L_{\mathbf{X}}$ to denote the real part of the above functional, that is, $L_{\mathbf{X}}$ denotes the linear form on $\mathbb{R}\langle\mathbf{x}\rangle$ defined by

$$L_{\mathbf{X}}(p) = \text{Re}(\text{Tr}(p(X_1, \ldots, X_n))) \quad \text{for} \quad p \in \mathbb{R}\langle\mathbf{x}\rangle. \tag{5}$$

Observe that $L_{\mathbf{X}}$ too is a symmetric tracial positive linear functional on $\mathbb{R}\langle\mathbf{x}\rangle$. Moreover, $L_{\mathbf{X}}$ is nonnegative on $\mathscr{M}(S)$ if the matrix tuple $\mathbf{X}$ is taken from the *matrix positivity domain* $\mathscr{D}(S)$ associated to the finite set $S \subseteq \text{Sym}\,\mathbb{R}\langle\mathbf{x}\rangle$, defined as

$$\mathscr{D}(S) = \bigcup_{d \geq 1} \left\{ \mathbf{X} = (X_1, \ldots, X_n) \in (\mathbf{H}^d)^n : g(\mathbf{X}) \succeq 0 \text{ for } g \in S \right\}. \tag{6}$$

Similarly, the linear functional $L_{\mathbf{X}}$ is zero on $\mathscr{I}(T)$ if the matrix tuple $\mathbf{X}$ is taken from the *matrix variety* $\mathscr{V}(T)$ associated to the finite set $T \subseteq \text{Sym}\,\mathbb{R}\langle\mathbf{x}\rangle$, defined as

$$\mathscr{V}(T) = \bigcup_{d \geq 1} \left\{ \mathbf{X} \in (\mathbf{H}^d)^n : h(\mathbf{X}) = 0 \text{ for all } h \in T \right\},$$

To discuss convergence properties of our lower bounds for matrix factorization ranks, we will need to consider infinite dimensional analogs of matrix algebras, namely

$C^*$-algebras admitting a tracial state. Let us introduce some basic notions we need about $C^*$-algebras; see, e.g., [9] for details. For our purposes, we define a $C^*$-*algebra* to be a norm closed $*$-subalgebra of the complex algebra $\mathscr{B}(\mathscr{H})$ of bounded operators on a complex Hilbert space $\mathscr{H}$. In particular, we have $\|a^*a\| = \|a\|^2$ for all elements $a$ in the algebra. Such an algebra $\mathscr{A}$ is said to be *unital* if it contains the identity operator (denoted 1). For instance, any full complex matrix algebra $\mathbb{C}^{d\times d}$ is a unital $C^*$-algebra. Moreover, by a fundamental result of Artin-Wedderburn, any finite dimensional $C^*$-algebra (as a vector space) is $*$-isomorphic to a direct sum $\bigoplus_{m=1}^{M} \mathbb{C}^{d_m \times d_m}$ of full complex matrix algebras [3,77]. In particular, any finite dimensional $C^*$-algebra is unital.

An element $b$ in a $C^*$-algebra $\mathscr{A}$ is called *positive*, denoted $b \succeq 0$, if it is of the form $b = a^*a$ for some $a \in \mathscr{A}$. For finite sets $S \subseteq \mathrm{Sym}\,\mathbb{R}\langle \mathbf{x}\rangle$ and $T \subseteq \mathbb{R}\langle \mathbf{x}\rangle$, the $C^*$-algebraic analogs of the matrix positivity domain and matrix variety are the sets

$$\mathscr{D}_{\mathscr{A}}(S) = \left\{\mathbf{X} = (X_1, \ldots, X_n) \in \mathscr{A}^n : X_i^* = X_i \text{ for } i \in [n],\ g(\mathbf{X}) \succeq 0 \text{ for } g \in S\right\},$$
$$\mathscr{V}_{\mathscr{A}}(T) = \left\{\mathbf{X} = (X_1, \ldots, X_n) \in \mathscr{A}^n : X_i^* = X_i \text{ for } i \in [n],\ h(\mathbf{X}) = 0 \text{ for } h \in T\right\}.$$

A *state* $\tau$ on a unital $C^*$-algebra $\mathscr{A}$ is a linear form on $\mathscr{A}$ that is *positive*, i.e., $\tau(a^*a) \geq 0$ for all $a \in \mathscr{A}$, and satisfies $\tau(1) = 1$. Since $\mathscr{A}$ is a complex algebra, every state $\tau$ is Hermitian: $\tau(a) = \overline{\tau(a^*)}$ for all $a \in \mathscr{A}$. We say that that a state is *tracial* if $\tau(ab) = \tau(ba)$ for all $a, b \in \mathscr{A}$ and *faithful* if $\tau(a^*a) = 0$ implies $a = 0$. A useful fact is that on a full matrix algebra $\mathbb{C}^{d \times d}$ the normalized matrix trace is the unique tracial state (see, e.g., [15]). Now, given a tuple $\mathbf{X} = (X_1, \ldots, X_n) \in \mathscr{A}^n$ in a $C^*$-algebra $\mathscr{A}$ with tracial state $\tau$, the second key example of a symmetric tracial positive linear functional on $\mathbb{R}\langle \mathbf{x}\rangle$ is given by the *trace evaluation map*, which we again denote by $L_{\mathbf{X}}$ and is defined by

$$L_{\mathbf{X}}(p) = \tau(p(X_1, \ldots, X_n)) \quad \text{for all} \quad p \in \mathbb{R}\langle \mathbf{x}\rangle.$$

### 1.4 Basic Approach

To explain the basic idea of how we obtain lower bounds for matrix factorization ranks, we consider the case of the completely positive semidefinite rank. Given a minimal factorization $A = (\mathrm{Tr}(X_i, X_j))$, with $d = \mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A)$ and $\mathbf{X} = (X_1, \ldots, X_n)$ in $(\mathrm{H}_+^d)^n$, consider the linear form $L_{\mathbf{X}}$ on $\mathbb{R}\langle \mathbf{x}\rangle$ as defined in (5):

$$L_{\mathbf{X}}(p) = \mathrm{Re}(\mathrm{Tr}(p(X_1, \ldots, X_n))) \quad \text{for} \quad p \in \mathbb{R}\langle \mathbf{x}\rangle.$$

Then, we have $A = (L_{\mathbf{X}}(x_i x_j))$ and $\mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A) = d = L_{\mathbf{X}}(1)$. To obtain lower bounds on $\mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A)$, we minimize $L(1)$ over a set of linear functionals $L$ that satisfy certain computationally tractable properties of $L_{\mathbf{X}}$. Note that this idea of minimizing $L(1)$ has recently been used in the works [59,74] in the commutative setting to derive a hierarchy of lower bounds converging to the nuclear norm of a symmetric tensor.

The above linear functional $L_{\mathbf{X}}$ is symmetric and tracial. Moreover, it satisfies some positivity conditions, since we have $L_{\mathbf{X}}(q) \geq 0$ whenever $q(\mathbf{X})$ is positive semidefinite. It follows that $L_{\mathbf{X}}(p^*p) \geq 0$ for all $p \in \mathbb{R}\langle\mathbf{x}\rangle$ and, as we explain later, $L_{\mathbf{X}}$ satisfies the *localizing* conditions $L_{\mathbf{X}}(p^*(\sqrt{A_{ii}}x_i - x_i^2)p) \geq 0$ for all $p$ and $i$. Truncating the linear form yields the following hierarchy of lower bounds:

$$\xi_t^{\mathrm{cpsd}}(A) = \min\Big\{L(1) : L \in \mathbb{R}\langle x_1, \ldots, x_n\rangle_{2t}^* \text{ tracial and symmetric,}$$
$$L(x_i x_j) = A_{ij} \quad \text{for} \quad i, j \in [n],$$
$$L \geq 0 \quad \text{on} \quad \mathscr{M}_{2t}\big(\{\sqrt{A_{11}}x_1 - x_1^2, \ldots, \sqrt{A_{nn}}x_n - x_n^2\}\big)\Big\}.$$

The bound $\xi_t^{\mathrm{cpsd}}(A)$ is computationally tractable (for small $t$). Indeed, as was explained in Sect. 1.3, the localizing constraint "$L \geq 0$ on $\mathscr{M}_{2t}(S)$" can be enforced by requiring certain matrices, whose entries are determined by $L$, to be positive semidefinite. This makes the problem defining $\xi_t^{\mathrm{cpsd}}(A)$ into a semidefinite program. The localizing conditions ensure the Archimedean property of the quadratic module, which permits to show certain convergence properties of the bounds $\xi_t^{\mathrm{cpsd}}(A)$.

The above approach extends naturally to the other matrix factorization ranks, using the following two basic ideas. First, since the cp-rank and the nonnegative rank deal with factorizations by *diagonal* matrices, we use linear functionals acting on classical *commutative* polynomials. Second, the *asymmetric* factorization ranks (psd-rank and nonnegative rank) can be seen as analogs of the symmetric ranks in the *partial matrix* setting, where we know only the values of $L$ on the quadratic monomials corresponding to entries in the off-diagonal blocks (this will require scaling of the factors in order to be able to define localizing constraints ensuring the Archimedean property). A main advantage of our approach is that it applies to all four matrix factorization ranks, after easy suitable adaptations.

## 1.5 Connection to Polynomial Optimization

In classical polynomial optimization, the problem is to find the global minimum of a commutative polynomial $f$ over a semialgebraic set of the form

$$D(S) = \{x \in \mathbb{R}^n : g(x) \geq 0 \text{ for } g \in S\},$$

where $S \subseteq \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \ldots, x_n]$ is a finite set of polynomials.[1] Tracial polynomial optimization is a noncommutative analog, where the problem is to minimize the normalized trace $\mathrm{tr}(f(\mathbf{X}))$ of a symmetric polynomial $f$ over a matrix positivity domain $\mathscr{D}(S)$ where $S \subseteq \mathrm{Sym}\,\mathbb{R}\langle\mathbf{x}\rangle$ is a finite set of symmetric polynomials.[2] Notice that the distinguishing feature here is the dimension independence: the optimization is over all possible matrix sizes. Perhaps counterintuitively, in this paper, we use techniques

---

[1] Here, and throughout the paper, we use $[\mathbf{x}]$ as the commutative analog of $\langle\mathbf{x}\rangle$.

[2] In fact, one could consider optimization over $\mathscr{D}(S) \cap \mathscr{V}(T)$ for some finite set $T \subseteq \mathbb{R}\langle\mathbf{x}\rangle$, the results below still hold in that setting, see Appendix A.

similar to those used for the tracial polynomial optimization problem to compute lower bounds on factorization dimensions.

For classical polynomial optimization Lasserre [46] and Parrilo [60] have proposed hierarchies of semidefinite programming relaxations based on the theory of moments and the dual theory of sums of squares polynomials. These can be used to compute successively better lower bounds converging to the global minimum (under the Archimedean condition). This approach has been used in a wide range of applications and there is an extensive literature (see, e.g., [1,47,49]). Most relevant to this work, it is used in [48] to design conic approximations of the completely positive cone and in [58] to check membership in the completely positive cone. This approach has also been extended to the noncommutative setting, first to the eigenvalue optimization problem [57,61] (which will not play a role in this paper), and later to tracial optimization [14,43].

For our paper, the moment formulation of the lower bounds is most relevant: For all $t \in \mathbb{N} \cup \{\infty\}$, we can define the bounds

$$f_t = \inf\{L(f) : L \in \mathbb{R}[\mathbf{x}]_{2t}^*, \ L(1) = 1, \ L \geq 0 \text{ on } \mathcal{M}_{2t}(S)\},$$
$$f_t^{\mathrm{tr}} = \inf\{L(f) : L \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}^* \text{ tracial and symmetric}, \ L(1) = 1, \ L \geq 0 \text{ on } \mathcal{M}_{2t}(S)\},$$

where $f_t$ (resp., $f_t^{\mathrm{tr}}$) lower bounds the (tracial) polynomial optimization problem.

The connection between the parameters $\xi_t^{\mathrm{cpsd}}(A)$ and $f_t^{\mathrm{tr}}$ is now clear: in the former we do not have the normalization property "$L(1) = 1$" but we do have the additional affine constraints "$L(x_i x_j) = A_{ij}$". This close relation to (tracial) polynomial optimization allows us to use that theory to understand the convergence properties of our bounds. Since throughout the paper we use (proof) techniques from (tracial) polynomial optimization, we will state the main convergence results we need, with full proofs, in Appendix A. Moreover, we give all proofs from the "moment side", which is most relevant to our treatment. Below we give a short summary of the convergence results for the hierarchies $\{f_t\}$ and $\{f_t^{\mathrm{tr}}\}$ that are relevant to our paper. We refer to Appendix A.3 for details.

Under the condition that $\mathcal{M}(S)$ is Archimedean, we have asymptotic convergence: $f_t \to f_\infty$ and $f_t^{\mathrm{tr}} \to f_\infty^{\mathrm{tr}}$ as $t \to \infty$. In the commutative setting, one can moreover show that $f_\infty$ is equal to the global minimum of $f$ over the set $D(S)$. However, in the noncommutative setting, the parameter $f_\infty^{\mathrm{tr}}$ is in general not equal to the minimum of $\mathrm{tr}(f(\mathbf{X}))$ over $\mathbf{X} \in \mathscr{D}(S)$. Instead, we need to consider the $C^*$-algebraic version of the tracial polynomial optimization problem: one can show that

$$f_\infty^{\mathrm{tr}} = \inf\{\tau(f(\mathbf{X})) : \mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S), \ \mathscr{A} \text{ is a unital } C^* \text{-algebra with tracial state } \tau\}.$$

An important additional convergence result holds under flatness. If the program defining the bound $f_t$ (resp., $f_t^{\mathrm{tr}}$) admits a sufficiently flat optimal solution, then equality holds: $f_t = f_\infty$ (resp., $f_t^{\mathrm{tr}} = f_\infty^{\mathrm{tr}}$). Moreover, in this case, the parameter $f_t^{\mathrm{tr}}$ is equal to the minimum value of $\mathrm{tr}(f(\mathbf{X}))$ over the matrix positivity domain $\mathscr{D}(S)$.

## 2 Lower Bounds on the Completely Positive Semidefinite Rank

Let $A$ be a completely positive semidefinite $n \times n$ matrix. For $t \in \mathbb{N} \cup \{\infty\}$, we consider the following semidefinite program, which, as we see below, lower bounds the complex completely positive semidefinite rank of $A$:

$$\xi_t^{\text{cpsd}}(A) = \min\{L(1) : L \in \mathbb{R}\langle x_1, \ldots, x_n \rangle_{2t}^* \text{ tracial and symmetric,}$$
$$L(x_i x_j) = A_{ij} \quad \text{for} \quad i, j \in [n],$$
$$L \geq 0 \quad \text{on} \quad \mathcal{M}_{2t}(S_A^{\text{cpsd}})\},$$

where we set

$$S_A^{\text{cpsd}} = \{\sqrt{A_{11}}x_1 - x_1^2, \ldots, \sqrt{A_{nn}}x_n - x_n^2\}. \tag{7}$$

Additionally, define the parameter $\xi_*^{\text{cpsd}}(A)$, obtained by adding the rank constraint $\text{rank}(M(L)) < \infty$ to the program defining $\xi_\infty^{\text{cpsd}}(A)$, where we consider the infimum instead of the minimum since we do not know whether the infimum is always attained. (In Proposition 1 we show the infimum is attained in $\xi_t^{\text{cpsd}}(A)$ for $t \in \mathbb{N} \cup \{\infty\}$). This gives a hierarchy of monotone nondecreasing lower bounds on the completely positive semidefinite rank:

$$\xi_1^{\text{cpsd}}(A) \leq \ldots \leq \xi_t^{\text{cpsd}}(A) \leq \ldots \leq \xi_\infty^{\text{cpsd}}(A) \leq \xi_*^{\text{cpsd}}(A) \leq \text{cpsd-rank}_{\mathbb{C}}(A).$$

The inequality $\xi_\infty^{\text{cpsd}}(A) \leq \xi_*^{\text{cpsd}}(A)$ is clear and monotonicity as well: If $L$ is feasible for $\xi_k^{\text{cpsd}}(A)$ with $t \leq k \leq \infty$, then its restriction to $\mathbb{R}\langle \mathbf{x} \rangle_{2t}$ is feasible for $\xi_t^{\text{cpsd}}(A)$.

The following notion of *localizing* polynomials will be useful. A set $S \subseteq \mathbb{R}\langle \mathbf{x} \rangle$ is said to be *localizing* at a matrix tuple $\mathbf{X}$ if $\mathbf{X} \in \mathscr{D}(S)$ (i.e., $g(\mathbf{X}) \succeq 0$ for all $g \in S$) and we say that $S$ is *localizing for $A$* if $S$ is localizing at some factorization $\mathbf{X} \in (\text{H}_+^d)^n$ of $A$ with $d = \text{cpsd-rank}_{\mathbb{C}}(A)$. The set $S_A^{\text{cpsd}}$ as defined in (7) is localizing for $A$, and, in fact, it is localizing at *any* factorization $\mathbf{X}$ of $A$ by Hermitian positive semidefinite matrices. Indeed, since

$$A_{ii} = \text{Tr}(X_i^2) \geq \lambda_{\max}(X_i^2) = \lambda_{\max}(X_i)^2$$

we have $\sqrt{A_{ii}}X_i - X_i^2 \succeq 0$ for all $i \in [n]$.

We can now use this to show the inequality $\xi_*^{\text{cpsd}}(A) \leq \text{cpsd-rank}_{\mathbb{C}}(A)$. For this set $d = \text{cpsd-rank}_{\mathbb{C}}(A)$, let $\mathbf{X} \in (\text{H}_+^d)^n$ be a Gram factorization of $A$, and consider the linear form $L_{\mathbf{X}} \in \mathbb{R}\langle \mathbf{x} \rangle^*$ defined by

$$L_{\mathbf{X}}(p) = \text{Re}(\text{Tr}(p(\mathbf{X}))) \quad \text{for all} \quad p \in \mathbb{R}\langle \mathbf{x} \rangle.$$

By construction $L_{\mathbf{X}}$ is symmetric and tracial, and we have $A = (L(x_i x_j))$. Moreover, since the set of polynomials $S_A^{\text{cpsd}}$ is localizing for $A$, the linear form $L_{\mathbf{X}}$ is nonnegative on $\mathcal{M}(S_A^{\text{cpsd}})$. Finally, we have $\text{rank}(M(L_{\mathbf{X}})) < \infty$, since the algebra generated by

$X_1, \ldots, X_n$ is finite dimensional. Hence, $L_{\mathbf{X}}$ is feasible for $\xi_*^{\text{cpsd}}(A)$ with $L_{\mathbf{X}}(1) = d$, which shows $\xi_*^{\text{cpsd}}(A) \leq \text{cpsd-rank}_{\mathbb{C}}(A)$.

The inclusions in (8) below show the quadratic module $\mathscr{M}(S_A^{\text{cpsd}})$ is Archimedean (recall the definition in (3)). Moreover, although there are other possible choices for the localizing polynomials to use in $S_A^{\text{cpsd}}$, these inclusions also show that the choice made in (7) leads to the largest truncated quadratic module and thus to the best bound. For any scalar $c > 0$, we have the inclusions

$$\mathscr{M}_{2t}(x, c - x) \subseteq \mathscr{M}_{2t}(x, c^2 - x^2) \subseteq \mathscr{M}_{2t}(cx - x^2) \subseteq \mathscr{M}_{2t+2}(x, c - x), \quad (8)$$

which hold in light of the following identities:

$$c - x = \left((c - x)^2 + c^2 - x^2\right)/(2c), \quad (9)$$
$$c^2 - x^2 = (c - x)^2 + 2(cx - x^2), \quad (10)$$
$$cx - x^2 = \left((c - x)x(c - x) + x(c - x)x\right)/c, \quad (11)$$
$$x = \left((cx - x^2) + x^2\right)/c. \quad (12)$$

In the rest of this section, we investigate properties of the hierarchy $\{\xi_t^{\text{cpsd}}(A)\}$ as well as some variations on it. We discuss convergence properties, asymptotically and under flatness, and we give another formulation for the parameter $\xi_*^{\text{cpsd}}(A)$. Moreover, as the inequality $\xi_*^{\text{cpsd}}(A) \leq \text{cpsd-rank}_{\mathbb{C}}(A)$ is typically strict, we present an approach to strengthen the bounds in order to go beyond $\xi_*^{\text{cpsd}}(A)$. Then, we propose some techniques to simplify the computation of the bounds, and we illustrate the behavior of the bounds on some examples.

## 2.1 The Parameters $\xi_\infty^{\text{cpsd}}(A)$ and $\xi_*^{\text{cpsd}}(A)$

In this section, we consider convergence properties of the hierarchy $\xi_t^{\text{cpsd}}(\cdot)$, both asymptotically and under flatness. We also give equivalent reformulations of the limiting parameters $\xi_\infty^{\text{cpsd}}(A)$ and $\xi_*^{\text{cpsd}}(A)$ in terms of $C^*$-algebras with a tracial state, which we will use in Sects. 2.3–2.4 to show properties of these parameters.

**Proposition 1** *Let $A \in \text{CS}_+^n$. For $t \in \mathbb{N} \cup \{\infty\}$ the optimum in $\xi_t^{\text{cpsd}}(A)$ is attained, and*

$$\lim_{t \to \infty} \xi_t^{\text{cpsd}}(A) = \xi_\infty^{\text{cpsd}}(A).$$

*Moreover, $\xi_\infty^{\text{cpsd}}(A)$ is equal to the smallest scalar $\alpha \geq 0$ for which there exists a unital $C^*$-algebra $\mathscr{A}$ with tracial state $\tau$ and $(X_1, \ldots, X_n) \in \mathscr{D}_{\mathscr{A}}(S_A^{\text{cpsd}})$ such that $A = \alpha \cdot (\tau(X_i X_j))$.*

**Proof** The sequence $(\xi_t^{\text{cpsd}}(A))_t$ is monotonically nondecreasing and upper bounded by $\xi_\infty^{\text{cpsd}}(A) < \infty$, which implies its limit exists and is at most $\xi_\infty^{\text{cpsd}}(A)$.

As $\xi_t^{\mathrm{cpsd}}(A) \leq \xi_\infty^{\mathrm{cpsd}}(A)$, we may add the redundant constraint $L(1) \leq \xi_\infty^{\mathrm{cpsd}}(A)$ to the problem $\xi_t^{\mathrm{cpsd}}(A)$ for every $t \in \mathbb{N}$. By (10), we have $\mathrm{Tr}(A) - \sum_i x_i^2 \in \mathscr{M}_2(S_A^{\mathrm{cpsd}})$. Hence, using the result of Lemma 13, the feasible region of $\xi_t^{\mathrm{cpsd}}(A)$ is compact, and thus, it has an optimal solution $L_t$. Again by Lemma 13, the sequence $(L_t)$ has a pointwise converging subsequence with limit $L \in \mathbb{R}\langle \mathbf{x} \rangle^*$. This pointwise limit $L$ is symmetric, tracial, satisfies $(L(x_i x_j)) = A$, and is nonnegative on $\mathscr{M}(S_A^{\mathrm{cpsd}})$. Hence, $L$ is feasible for $\xi_\infty^{\mathrm{cpsd}}(A)$. This implies that $L$ is optimal for $\xi_\infty^{\mathrm{cpsd}}(A)$ and we have $\lim_{t \to \infty} \xi_t^{\mathrm{cpsd}}(A) = \xi_\infty^{\mathrm{cpsd}}(A)$.

The reformulation of $\xi_\infty^{\mathrm{cpsd}}(A)$ in terms of $C^*$-algebras with a tracial state follows directly using Theorem 1.    □

Next, we give some equivalent reformulations for the parameter $\xi_*^{\mathrm{cpsd}}(A)$, which follow as a direct application of Theorem 2. In general, we do not know whether the infimum in $\xi_*^{\mathrm{cpsd}}(A)$ is attained. However, as a direct application of Corollary 1, we see that this infimum is attained if there is an integer $t \in \mathbb{N}$ for which $\xi_t^{\mathrm{cpsd}}(A)$ admits a flat optimal solution.

**Proposition 2** *Let $A \in \mathrm{CS}_+^n$. The parameter $\xi_*^{\mathrm{cpsd}}(A)$ is given by the infimum of $L(1)$ taken over all conic combinations $L$ of trace evaluations at elements in $\mathscr{D}_{\mathscr{A}}(S_A^{\mathrm{cpsd}})$ for which $A = (L(x_i x_j))$. The parameter $\xi_*^{\mathrm{cpsd}}(A)$ is also equal to the infimum over all $\alpha \geq 0$ for which there exist a finite dimensional $C^*$-algebra $\mathscr{A}$ with tracial state $\tau$ and $(X_1, \ldots, X_n) \in \mathscr{D}_{\mathscr{A}}(S_A^{\mathrm{cpsd}})$ such that $A = \alpha \cdot (\tau(X_i X_j))$.*

*In addition, if $\xi_t^{\mathrm{cpsd}}(A)$ admits a flat optimal solution, then $\xi_t^{\mathrm{cpsd}}(A) = \xi_*^{\mathrm{cpsd}}(A)$.*

Next we show a formulation for $\xi_*^{\mathrm{cpsd}}(A)$ in terms of factorization by block-diagonal matrices, which helps explain why the inequality $\xi_*^{\mathrm{cpsd}}(A) \leq \mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A)$ is typically strict. Here $\|\cdot\|$ is the operator norm, so that $\|X\| = \lambda_{\max}(X)$ for $X \succeq 0$.

**Proposition 3** *For $A \in \mathrm{CS}_+^n$ we have*

$$\xi_*^{\mathrm{cpsd}}(A) = \inf\left\{ \sum_{m=1}^M d_m \cdot \max_{i \in [n]} \frac{\|X_i^m\|^2}{A_{ii}} : M \in \mathbb{N},\ d_1, \ldots, d_M \in \mathbb{N}, \right.$$

$$X_i^m \in \mathrm{H}_+^{d_m} \text{ for } i \in [n], m \in [M],$$

$$\left. A = \mathrm{Gram}\left( \oplus_{m=1}^M X_1^m, \ldots, \oplus_{m=1}^M X_n^m \right) \right\}.$$

$$(13)$$

*Note that using matrices from $\mathrm{S}_+^{d_m}$ instead of $\mathrm{H}_+^{d_m}$ does not change the optimal value.*

**Proof** The proof uses the formulation of $\xi_*^{\mathrm{cpsd}}(A)$ in terms of conic combinations of trace evaluations at matrix tuples in $\mathscr{D}(S_A^{\mathrm{cpsd}})$ as given in Proposition 2. We first show the inequality $\beta \leq \xi_*^{\mathrm{cpsd}}(A)$, where $\beta$ denotes the optimal value of the program in (13).

For this, assume $L \in \mathbb{R}\langle \mathbf{x} \rangle^*$ is a conic combination of trace evaluations at elements of $\mathscr{D}(S_A^{\text{cpsd}})$ such that $A = (L(x_i x_j))$. We will construct a feasible solution for (13) with objective value $L(1)$. The linear functional $L$ can be written as

$$L = \sum_{m=1}^{M} \lambda_m L_{\mathbf{Y}^m}, \text{ where } \lambda_m > 0 \text{ and } \mathbf{Y}^m = (Y_1^m, \ldots, Y_n^m) \in \mathscr{D}(S_A^{\text{cpsd}}) \text{ for } m \in [M].$$

Let $d_m$ denote the size of the matrices $Y_1^m, \ldots, Y_n^m$, so that $L(1) = \sum_m \lambda_m d_m$. Since $\mathbf{Y}^m \in \mathscr{D}(S_A^{\text{cpsd}})$, we have $Y_i^m \succeq 0$ and $A_{ii} I - (Y_i^m)^2 \succeq 0$ by identities (10) and (12). This implies $\|Y_i^m\|^2 \leq A_{ii}$ for all $i \in [n]$ and $m \in [M]$. Define $\mathbf{X}^m = \sqrt{\lambda_m}\, \mathbf{Y}^m$. Then, $L(x_i x_j) = \sum_m \text{Tr}(X_i^m X_j^m)$, so that the matrices $\oplus_m X_1^m, \ldots, \oplus_m X_n^m$ form a Gram decomposition of $A$. This gives a feasible solution to (13) with value

$$\sum_{m=1}^{M} d_m \cdot \max_{i \in [n]} \frac{\|X_i^m\|^2}{A_{ii}} = \sum_{m=1}^{M} d_m \lambda_m \max_{i \in [n]} \frac{\|Y_i^m\|^2}{A_{ii}} \leq \sum_{m=1}^{M} d_m \lambda_m = L(1),$$

which shows $\beta \leq L(1)$, and hence $\beta \leq \xi_*^{\text{cpsd}}(A)$.

For the other direction, we assume

$$A = \text{Gram}\big(\oplus_{m=1}^{M} X_1^m, \ldots, \oplus_{m=1}^{M} X_n^m\big), \quad X_1^m, \ldots, X_n^m \in S_+^{d_m} \text{ for } m \in [M].$$

Set $\lambda_m = \max_{i \in [n]} \|X_i^m\|^2 / A_{ii}$, and define the linear form $L$ by

$$L = \sum_{m=1}^{M} \lambda_m L_{\mathbf{Y}^m}, \quad \text{where} \quad \mathbf{Y}^m = \mathbf{X}^m / \sqrt{\lambda_m} \quad \text{for all} \quad m \in [M].$$

We have $L(1) = \sum_m \lambda_m d_m$ and $A = (L(x_i x_j))$, and thus it suffices to show that each matrix tuple $\mathbf{Y}^m$ belongs to $\mathscr{D}(S_A^{\text{cpsd}})$. For this we observe that $\lambda_m A_{ii} \geq \|X_i^m\|^2$. Therefore $\lambda_m A_{ii} I \succeq (X_i^m)^2$, and thus $A_{ii} I \succeq (Y_i^m)^2$, which implies $\sqrt{A_{ii}} Y_i^m - (Y_i^m)^2 \succeq 0$. This shows $\xi_*^{\text{cpsd}}(A) \leq L(1) = \sum_m \lambda_m d_m$, and thus $\xi_*^{\text{cpsd}}(A) \leq \beta$.                                                                                     $\square$

We can say a bit more when the matrix $A$ lies on an extreme ray of the cone $\text{CS}_+^n$. In the formulation from Proposition 3, it suffices to restrict the minimization over factorizations of $A$ involving only one block. However, we know very little about the extreme rays of $\text{CS}_+^n$, also in view of the recent result that the cone is not closed for large $n$ [22,73].

**Proposition 4** *If $A$ lies on an extreme ray of the cone $\text{CS}_+^n$, then*

$$\xi_*^{\text{cpsd}}(A) = \inf \left\{ d \cdot \max_{i \in [n]} \frac{\|X_i\|^2}{A_{ii}} : d \in \mathbb{N}, X_1, \ldots, X_n \in \mathbb{H}_+^d, A = \text{Gram}(X_1, \ldots, X_n) \right\}.$$

*Moreover, if $\oplus_{m=1}^{M} X_1^m, \ldots, \oplus_{m=1}^{M} X_n^m$ is a Gram decomposition of $A$ providing an optimal solution to (13) and some block $X_i^m$ has rank 1, then $\xi_*^{\text{cpsd}}(A) = \text{cpsd-rank}_{\mathbb{C}}(A)$.*

**Proof** Let $\beta$ be the infimum in Proposition 4. The inequality $\xi_*^{\text{cpsd}}(A) \leq \beta$ follows from the reformulation of $\xi_*^{\text{cpsd}}(A)$ in Proposition 3. To show the reverse inequality we consider a solution $\oplus_{m=1}^M X_1^m, \ldots, \oplus_{m=1}^M X_n^m$ to (13), and set $\lambda_m = \max_i \|X_i^m\|^2 / A_{ii}$. We will show $\beta \leq \sum_m d_m \lambda_m$. For this define the matrices $A_m = \text{Gram}(X_1^m, \cdots, X_n^m)$, so that $A = \sum_m A_m$. As $A$ lies on an extreme ray of $\text{CS}_+^n$, we must have $A_m = \alpha_m A$ for some $\alpha_m > 0$ with $\sum_m \alpha_m = 1$. Hence, since

$$A = A_m / \alpha_m = \text{Gram}(X_1^m / \sqrt{\alpha_m}, \cdots, X_n^m / \sqrt{\alpha_m}),$$

we have $\beta \leq d_m \lambda_m / \alpha_m$ for all $m \in [M]$. It suffices now to use $\sum_m \alpha_m = 1$ to see that $\min_m d_m \lambda_m / \alpha_m \leq \sum_m d_m \lambda_m$. So we have shown $\beta \leq \min_m d_m \lambda_m / \alpha_m \leq \sum_m d_m \lambda_m$. This implies $\beta \leq \xi_*^{\text{cpsd}}(A)$, and thus equality holds.

Assume now that $\oplus_{m=1}^M X_1^m, \ldots, \oplus_{m=1}^M X_n^m$ is optimal to (13) and that there is a block $X_i^m$ of rank 1. By Proposition 3 we have $\sum_m d_m \lambda_m = \xi_*^{\text{cpsd}}(A)$. From the argument just made above it follows that

$$\xi_*^{\text{cpsd}}(A) = \min_m d_m \lambda_m / \alpha_m = \sum_m d_m \lambda_m.$$

As $\sum_m \alpha_m = 1$ this implies $d_m \lambda_m / \alpha_m = \min_m d_m \lambda_m / \alpha_m$ for all $m$; that is, all terms $d_m \lambda_m / \alpha_m$ take the same value $\xi_*^{\text{cpsd}}(A)$. By assumption, there exist some $m \in [M]$ and $i \in [n]$ for which $X_i^m$ has rank 1. Then $\|X_i^m\|^2 = \langle X_i^m, X_i^m \rangle$, which gives $\lambda_m = \alpha_m$, and thus $\xi_*^{\text{cpsd}}(A) = d_m$. On the other hand, $\text{cpsd-rank}_{\mathbb{C}}(A) \leq d_m$ since $(X_i^m / \sqrt{\alpha_m})_i$ forms a Gram decomposition of $A$, so equality $\xi_*^{\text{cpsd}}(A) = d_m = \text{cpsd-rank}_{\mathbb{C}}(A)$ holds. $\square$

## 2.2 Additional Localizing Constraints to Improve on $\xi_*^{\text{cpsd}}(A)$

In order to strengthen the bounds, we may require nonnegativity over a (truncated) quadratic module generated by a larger set of localizing polynomials for $A$. The following lemma gives one such approach.

**Lemma 1** *Let* $A \in \text{CS}_+^n$. *For* $v \in \mathbb{R}^n$ *and* $g_v = v^T A v - \left(\sum_{i=1}^n v_i x_i\right)^2$, *the set* $\{g_v\}$ *is localizing for every Gram factorization by Hermitian positive semidefinite matrices of* $A$ *(in particular,* $\{g_v\}$ *is localizing for* $A$*).*

**Proof** If $X_1, \ldots, X_n$ is a Gram decomposition of $A$ by Hermitian positive semidefinite matrices, then

$$v^T A v = \text{Tr}\left(\left(\sum_{i=1}^n v_i X_i\right)^2\right) \geq \lambda_{\max}\left(\left(\sum_{i=1}^n v_i X_i\right)^2\right),$$

hence $v^T A v I - (\sum_{i=1}^n v_i X_i)^2 \succeq 0$. $\square$

Given a set $V \subseteq \mathbb{R}^n$, we consider the larger set

$$S_{A,V}^{\text{cpsd}} = S_A^{\text{cpsd}} \cup \{g_v : v \in V\}$$

of localizing polynomials for $A$. For $t \in \mathbb{N} \cup \{\infty, *\}$, denote by $\xi_{t,V}^{\text{cpsd}}(A)$ the parameter obtained by replacing in $\xi_t^{\text{cpsd}}(A)$ the nonnegativity constraint on $\mathscr{M}_{2t}(S_A^{\text{cpsd}})$ by nonnegativity on the larger set $\mathscr{M}_{2t}(S_{A,V}^{\text{cpsd}})$. We have $\xi_{t,\emptyset}^{\text{cpsd}}(A) = \xi_t^{\text{cpsd}}(A)$ and

$$\xi_t^{\text{cpsd}}(A) \leq \xi_{t,V}^{\text{cpsd}}(A) \leq \text{cpsd-rank}_{\mathbb{C}}(A) \quad \text{for all} \quad V \subseteq \mathbb{R}^n.$$

By scaling invariance, we can add the above constraints for all $v \in \mathbb{R}^n$ by setting $V$ to be the unit sphere $\mathbb{S}^{n-1}$. Since $\mathbb{S}^{n-1}$ is a compact metric space, there exists a sequence $V_1 \subseteq V_2 \subseteq \ldots \subseteq \mathbb{S}^{n-1}$ of finite subsets such that $\bigcup_{k \geq 1} V_k$ is dense in $\mathbb{S}^{n-1}$. Each of the parameters $\xi_{t,V_k}^{\text{cpsd}}(A)$ involves finitely many localizing constraints, and, as we now show, they converge to the parameter $\xi_{t,\mathbb{S}^{n-1}}^{\text{cpsd}}(A)$.

**Proposition 5** *Consider a matrix $A \in \text{CS}_+^n$. For $t \in \{\infty, *\}$, we have*

$$\lim_{k \to \infty} \xi_{t,V_k}^{\text{cpsd}}(A) = \xi_{t,\mathbb{S}^{n-1}}^{\text{cpsd}}(A).$$

**Proof** Let $\varepsilon > 0$. Since $\bigcup_k V_k$ is dense in $\mathbb{S}^{n-1}$, there is an integer $k \geq 1$ so that for every $u \in \mathbb{S}^{n-1}$ there exists a vector $v \in V_k$ satisfying

$$\|u - v\|_1 \leq \frac{\varepsilon \lambda_{\min}(A)}{4\sqrt{n} \max_i A_{ii}} \quad \text{and} \quad \|u - v\|_2 \leq \frac{\varepsilon \lambda_{\min}(A)}{4\text{Tr}(A^2)^{1/2}}. \tag{14}$$

The above Propositions 1 and 2 have natural analogs for the programs $\xi_{t,V}^{\text{cpsd}}(A)$. These show that for $t = \infty$ ($t = *$) the parameter $\xi_{t,V_k}^{\text{cpsd}}(A)$ is the infimum over all $\alpha \geq 0$ for which there exist a (finite dimensional) unital $C^*$-algebra $\mathscr{A}$ with tracial state $\tau$ and $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S_{A,V_k}^{\text{cpsd}})$ such that $A = \alpha \cdot (\tau(X_i X_j))$.

Below we will show that $\mathbf{X}' = \sqrt{1 - \varepsilon} \mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S_{A,\mathbb{S}^{n-1}}^{\text{cpsd}})$. This implies that the linear form $L \in \mathbb{R}\langle \mathbf{x} \rangle^*$ defined by $L(p) = \alpha/(1-\varepsilon)\tau(p(\mathbf{X}'))$ is feasible for $\xi_{t,\mathbb{S}^{n-1}}^{\text{cpsd}}(A)$ with objective value $L(1) = \alpha/(1 - \varepsilon)$. This shows

$$\xi_{t,\mathbb{S}^{n-1}}^{\text{cpsd}}(A) \leq \frac{1}{1-\varepsilon} \xi_{t,V_k}^{\text{cpsd}}(A) \leq \frac{1}{1-\varepsilon} \lim_{k \to \infty} \xi_{t,V_k}^{\text{cpsd}}(A).$$

Since $\varepsilon > 0$ was arbitrary, letting $\varepsilon$ tend to 0 completes the proof.

We now show $\mathbf{X}' = \sqrt{1 - \varepsilon} \mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S_{A,\mathbb{S}^{n-1}}^{\text{cpsd}})$. For this consider the map

$$f_{\mathbf{X}} \colon \mathbb{S}^{n-1} \to \mathbb{R}, \ v \mapsto \left\| \sum_{i=1}^n v_i X_i \right\|^2,$$

where $\|\cdot\|$ denotes the $C^*$-algebra norm of $\mathscr{A}$. For $\alpha \in \mathbb{R}$ and $a \in \mathscr{A}$ with $a^* = a$, we have $\alpha \geq \|a\|$ if and only if $\alpha - a \succeq 0$ in $\mathscr{A}$, or, equivalently, $\alpha^2 - a^2 \succeq 0$ in $\mathscr{A}$. Since $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S_{A,V_k}^{\mathrm{cpsd}})$ we have $v^\mathsf{T} A v - f_{\mathbf{X}}(v) \geq 0$ for all $v \in V_k$, and hence

$$v^\mathsf{T} A v - f_{\mathbf{X}'}(v) = v^\mathsf{T} A v \left(1 - (1-\varepsilon)\frac{f_{\mathbf{X}}(v)}{v^\mathsf{T} A v}\right) \geq v^\mathsf{T} A v \left(1 - (1-\varepsilon)\right)$$

$$= \varepsilon v^\mathsf{T} A v \geq \varepsilon \lambda_{\min}(A).$$

Let $u \in \mathbb{S}^{n-1}$ and let $v \in V_k$ be such that (14) holds. Using Cauchy-Schwarz we have

$$|u^\mathsf{T} A u - v^\mathsf{T} A v| = |(u-v)^\mathsf{T} A(u+v)| = |\langle A, (u-v)(u+v)^\mathsf{T}\rangle|$$

$$\leq \sqrt{\mathrm{Tr}(A^2)}\sqrt{\mathrm{Tr}((u+v)(u-v)^\mathsf{T}(u-v)(u+v)^\mathsf{T})}$$

$$\leq \sqrt{\mathrm{Tr}(A^2)}\|u-v\|_2\|u+v\|_2 \leq 2\sqrt{\mathrm{Tr}(A^2)}\|u-v\|_2$$

$$\leq 2\sqrt{\mathrm{Tr}(A^2)}\frac{\varepsilon\lambda_{\min}(A)}{4\sqrt{\mathrm{Tr}(A^2)}} = \frac{\varepsilon\lambda_{\min}(A)}{2}.$$

Since $\sqrt{A_{ii}}X_i - X_i^2$ is positive in $\mathscr{A}$, we have that $\sqrt{A_{ii}} - X_i$ is positive in $\mathscr{A}$ by (9) and (10), which implies $\|X_i\| \leq \sqrt{A_{ii}}$. By the reverse triangle inequality, we then have

$$|f_{\mathbf{X}'}(u) - f_{\mathbf{X}'}(v)| = \left|\left\|\sum_{i=1}^n u_i X_i'\right\| - \left\|\sum_{i=1}^n v_i X_i'\right\|\right| \left(\left\|\sum_{i=1}^n u_i X_i'\right\| + \left\|\sum_{i=1}^n v_i X_i'\right\|\right)$$

$$\leq \left\|\sum_{i=1}^n (v_i - u_i)X_i'\right\| 2\sqrt{n}\max_i\sqrt{A_{ii}}$$

$$\leq \left(\sum_{i=1}^n |v_i - u_i|\|X_i'\|\right) 2\sqrt{n}\max_i\sqrt{A_{ii}}$$

$$\leq \|u-v\|_1 2\sqrt{n}\max_i A_{ii}$$

$$\leq \frac{\varepsilon\lambda_{\min}(A)}{4\sqrt{n}\max_i A_{ii}} 2\sqrt{n}\max_i A_{ii} = \frac{\varepsilon\lambda_{\min}(A)}{2}.$$

Combining the above inequalities we obtain that $u^\mathsf{T} A u - f_{\mathbf{X}'}(u) \geq 0$ for all $\mathbb{S}^{n-1}$, and hence $u^\mathsf{T} A u - \left(\sum_{i=1}^n u_i X_i'\right)^2$ is positive in $\mathscr{A}$. Thus, we have $\mathbf{X}' \in \mathscr{D}_{\mathscr{A}}(S_{A,\mathbb{S}^{n-1}}^{\mathrm{cpsd}})$. □

We now discuss two examples where the bounds $\xi_{*,V}^{\mathrm{cpsd}}(A)$ go beyond $\xi_*^{\mathrm{cpsd}}(A)$.

**Example 1** Consider the matrix

$$A = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} = \mathrm{Gram}\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}\right), \tag{15}$$

with $\text{cpsd-rank}_{\mathbb{C}}(A) = 2$. We can also write $A = \text{Gram}(Y_1, Y_2)$, where

$$Y_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Y_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

With $X_i = \sqrt{2}\, Y_i$ we have $I - X_i^2 \succeq 0$ for $i = 1, 2$. Hence the linear form $L = L_{\mathbf{X}}/2$ is feasible for $\xi_*^{\text{cpsd}}(A)$, which shows that $\xi_*^{\text{cpsd}}(A) \leq L(1) = 3/2$. In fact, this form $L$ gives an optimal flat solution to $\xi_2^{\text{cpsd}}(A)$, as we can check using a semidefinite programming solver, so $\xi_*^{\text{cpsd}}(A) = 3/2$. In passing, we observe that $\xi_1^{\text{cpsd}}(A) = 4/3$, which coincides with the analytic lower bound (18) (see also Lemma 6 below).

For $e = (1, 1) \in \mathbb{R}^2$ and $V = \{e\}$, this form $L$ is not feasible for $\xi_{*,V}^{\text{cpsd}}(A)$, because for the polynomial $p = 1 - 3x_1 - 3x_2$ we have $L(p^* g_e p) = -9/2 < 0$. This means that the localizing constraint $L(p^* g_e p) \geq 0$ is not redundant: For $t \geq 2$ it cuts off part of the feasibility region of $\xi_t^{\text{cpsd}}(A)$. Indeed, using a semidefinite programming solver, we find an optimal flat solution of $\xi_{3,V}^{\text{cpsd}}(A)$ with objective value $(5 - \sqrt{3})/2 \approx 1.633$, hence

$$\xi_{*,V}^{\text{cpsd}}(A) = (5 - \sqrt{3})/2 > 3/2 = \xi_*^{\text{cpsd}}(A).$$

**Example 2** Consider the symmetric circulant matrices

$$M(\alpha) = \begin{pmatrix} 1 & \alpha & 0 & 0 & \alpha \\ \alpha & 1 & \alpha & 0 & 0 \\ 0 & \alpha & 1 & \alpha & 0 \\ 0 & 0 & \alpha & 1 & \alpha \\ \alpha & 0 & 0 & \alpha & 1 \end{pmatrix} \quad \text{for} \quad \alpha \in \mathbb{R}.$$

For $0 \leq \alpha \leq 1/2$, we have $M(\alpha) \in \text{CS}_+^5$ with $\text{cpsd-rank}_{\mathbb{C}}(M(\alpha)) \leq 5$. To see this, we set $\beta = (1 + \sqrt{1 - 4\alpha^2})/2$ and observe that the matrices

$$X_i = \text{Diag}(\sqrt{\beta}\, e_i + \sqrt{1 - \beta}\, e_{i+1}) \in S_+^5, \quad i \in [5], \quad (\text{with } e_6 := e_1),$$

form a factorization of $M(\alpha)$. As $M(\alpha)$ is supported by a cycle, we have $M(\alpha) \in \text{CS}_+^5$ if and only if $M(\alpha) \in \text{CP}^5$ [50]. Thus, $M(\alpha) \in \text{CS}_+^5$ if and only if $0 \leq \alpha \leq 1/2$.

By using its formulation in Proposition 3, we can use the above factorization to derive the inequality $\xi_*^{\text{cpsd}}(M(1/2)) \leq 5/2$. However, using a semidefinite programming solver, we see that

$$\xi_{2,V}^{\text{cpsd}}(M(1/2)) = 5,$$

where $V$ is the set containing the vector $(1, -1, 1, -1, 1)$ and its cyclic shifts. Hence, the bound $\xi_{2,V}^{\text{cpsd}}(M(1/2))$ is tight: It certifies $\text{cpsd-rank}_{\mathbb{C}}(M(1/2)) = 5$, while the

other known bounds, the rank bound $\sqrt{\operatorname{rank}(A)}$ and the analytic bound (18), only give cpsd-rank$_{\mathbb{C}}(A) \geq 3$.

We now observe that there exist $0 < \varepsilon, \delta < 1/2$ such that cpsd-rank$_{\mathbb{C}}(M(\alpha)) = 5$ for all $\alpha \in [0, \varepsilon] \cup [\delta, 1/2]$. Indeed, this follows from the fact that $\xi_1^{\mathrm{cpsd}}(M(0)) = 5$ (by Lemma 6), the above result that $\xi_{2,V}^{\mathrm{cpsd}}(M(1/2)) = 5$, and the lower semicontinuity of $\alpha \mapsto \xi_{2,V}^{\mathrm{cpsd}}(M(\alpha))$, which is shown in Lemma 7 below.

As the matrices $M(\alpha)$ are nonsingular, the above factorization shows that their cp-rank is equal to 5 for all $\alpha \in [0, 1/2]$; whether they all have cpsd-rank equal to 5 is not known.

### 2.3 Boosting the Bounds

In this section, we propose some additional constraints that can be added to strengthen the bounds $\xi_{t,V}^{\mathrm{cpsd}}(A)$ for finite $t$. These constraints may shrink the feasibility region of $\xi_{t,V}^{\mathrm{cpsd}}(A)$ for $t \in \mathbb{N}$, but they are redundant for $t \in \{\infty, *\}$. The latter is shown using the reformulation of the parameters $\xi_{\infty,V}^{\mathrm{cpsd}}(A)$ and $\xi_{*,V}^{\mathrm{cpsd}}(A)$ in terms of $C^*$-algebras.

We first mention how to construct localizing constraints of "bilinear type", inspired by the work of Berta, Fawzi and Scholz [7]. Note that as for localizing constraints, these bilinear constraints can be modeled as semidefinite constraints.

**Lemma 2** *Let $A \in \mathrm{CS}_+^n$, $t \in \mathbb{N} \cup \{\infty, *\}$, and let $\{g, g'\}$ be localizing for $A$. If we add the constraints*

$$L(p^* g p g') \geq 0 \quad for \quad p \in \mathbb{R}\langle \mathbf{x} \rangle \quad with \quad \deg(p^* g p g') \leq 2t \tag{16}$$

*to $\xi_{t,V}^{\mathrm{cpsd}}(A)$, then we still get a lower bound on* cpsd-rank$_{\mathbb{C}}(A)$. *However, the constraints (16) are redundant for $\xi_{\infty,V}^{\mathrm{cpsd}}(A)$ and $\xi_{*,V}^{\mathrm{cpsd}}(A)$ when $g, g' \in \mathscr{M}(S_{A,V}^{\mathrm{cpsd}})$.*

**Proof** Let $\mathbf{X} \in (\mathrm{H}_+^d)^n$ be a Gram decomposition of $A$, and let $L = L_{\mathbf{X}}$ be the real part of the trace evaluation at $\mathbf{X}$. Then, $p(\mathbf{X})^* g(\mathbf{X}) p(\mathbf{X}) \succeq 0$ and $g'(\mathbf{X}) \succeq 0$, and thus

$$L(p^* g p g') = \operatorname{Re}(\operatorname{Tr}(p(\mathbf{X})^* g(\mathbf{X}) p(\mathbf{X}) g'(\mathbf{X}))) \geq 0.$$

So by adding the constraints (16) we still get a lower bound on cpsd-rank$_{\mathbb{C}}(A)$.

To show that the constraints (16) are redundant for $\xi_{t,V}^{\mathrm{cpsd}}(A)$ and $\xi_{*,V}^{\mathrm{cpsd}}(A)$ when $g, g' \in \mathscr{M}(S_{A,V}^{\mathrm{cpsd}})$, we let $t \in \{\infty, *\}$ and assume $L$ is feasible for $\xi_{t,V}^{\mathrm{cpsd}}(A)$. By Theorem 1 there exist a unital $C^*$-algebra $\mathscr{A}$ with tracial state $\tau$ and $\mathbf{X} \in \mathscr{D}(S_{A,V}^{\mathrm{cpsd}})$ such that $L(p) = L(1)\tau(p(\mathbf{X}))$ for all $p \in \mathbb{R}\langle \mathbf{x} \rangle$. Since $g, g' \in \mathscr{M}(S_{A,V}^{\mathrm{cpsd}})$ we know that $g(\mathbf{X}), g'(\mathbf{X})$ are positive elements in $\mathscr{A}$, so $g(\mathbf{X}) = a^*a$ and $g'(\mathbf{X}) = b^*b$ for some $a, b \in \mathscr{A}$. Then, we have

$$L(p^* g p g) = L(1)\, \tau(p^*(\mathbf{X})\, g(\mathbf{X})\, p(\mathbf{X})\, g'(\mathbf{X}))$$
$$= L(1)\, \tau(p^*(\mathbf{X})\, a^*a\, p(\mathbf{X})\, b^*b)$$

$$= L(1)\,\tau((a\,p(\mathbf{X})\,b^*)^*a\,p(\mathbf{X})\,b^*) \geq 0,$$

where we use that $\tau$ is a positive tracial state on $\mathscr{A}$. $\qquad\square$

Second, we show how to use zero entries in $A$ and vectors in the kernel of $A$ to enforce new constraints on $\xi_{t,V}^{\mathrm{cpsd}}(A)$.

**Lemma 3** *Let* $A \in \mathrm{CS}_+^n$ *and* $t \in \mathbb{N} \cup \{\infty, *\}$. *If we add the constraint*

$$L = 0 \quad \text{on} \quad \mathscr{I}_{2t}\left(\left\{\sum_{i=1}^n v_i x_i : v \in \ker A\right\} \cup \left\{x_i x_j : A_{ij} = 0\right\}\right) \qquad (17)$$

*to* $\xi_{t,V}^{\mathrm{cpsd}}(A)$, *then we still get a lower bound on* cpsd-rank$_{\mathbb{C}}(A)$. *Moreover, these constraints are redundant for* $\xi_{\infty,V}^{\mathrm{cpsd}}(A)$ *and* $\xi_{*,V}^{\mathrm{cpsd}}(A)$.

**Proof** Let $\mathbf{X} \in (\mathrm{H}_+^d)^n$ be a Gram factorization of $A$ and let $L_{\mathbf{X}}$ be as in (5). If $Av = 0$, then $0 = v^{\mathsf{T}} A v = \mathrm{Tr}((\sum_{i=1}^n v_i X_i)^2)$ and thus $\sum_{i=1}^n v_i X_i = 0$. Hence $L_{\mathbf{X}}((\sum_{i=1}^n v_i x_i)p) = \mathrm{Re}(\mathrm{Tr}((\sum_{i=1}^n v_i X_i)p(\mathbf{X}))) = 0$. If $A_{ij} = 0$, then $\mathrm{Tr}(X_i X_j) = 0$, which implies $X_i X_j = 0$, since $X_i$ and $X_j$ are positive semidefinite. Hence, $L_{\mathbf{X}}(x_i x_j\,p) = \mathrm{Re}(\mathrm{Tr}(X_i X_j\,p(\mathbf{X}))) = 0$. Therefore, adding the constraints (17) still lower bounds cpsd-rank$_{\mathbb{C}}(A)$.

As in the proof of the previous lemma, if $t \in \{\infty, *\}$ and $L$ is feasible for $\xi_{t,V}^{\mathrm{cpsd}}(A)$ then, by Theorem 1, there exist a unital $C^*$-algebra $\mathscr{A}$ with tracial state $\tau$ and $\mathbf{X}$ in $\mathscr{D}(S_{A,V}^{\mathrm{cpsd}})$ such that $L(p) = L(1)\tau(p(\mathbf{X}))$ for all $p \in \mathbb{R}\langle\mathbf{x}\rangle$. Moreover, by Lemma 12, we may assume $\tau$ to be faithful. For a vector $v$ in the kernel of $A$, we have $0 = v^{\mathsf{T}} A v = L((\sum_i v_i x_i)^2) = L(1)\tau((\sum_i v_i X_i)^2)$, and hence, since $\tau$ is faithful, $\sum_i v_i X_i = 0$ in $\mathscr{A}$. It follows that $L(p(\sum_i v_i x_i)) = L(1)\tau(p(\mathbf{X})\,0) = 0$ for all $p \in \mathbb{R}\langle\mathbf{x}\rangle$. Analogously, if $A_{ij} = 0$, then $L(x_i x_j) = 0$ implies $\tau(X_i X_j) = 0$ and thus $X_i X_j = 0$, since $X_i, X_j$ are positive in $\mathscr{A}$ and $\tau$ is faithful. This implies $L(px_i x_j) = 0$ for all $p \in \mathbb{R}\langle\mathbf{x}\rangle$. This shows that the constraints (17) are redundant. $\qquad\square$

Note that the constraints $L(p\,(\sum_{i=1}^n v_i x_i)) = 0$ for $p \in \mathbb{R}\langle\mathbf{x}\rangle_t$, which are implied by (17), are in fact redundant: if $v \in \ker(A)$, then the vector obtained by extending $v$ with zeros belongs to $\ker(M_t(L))$, since $M_t(L) \succeq 0$. Also, for an implementation of $\xi_t^{\mathrm{cpsd}}(A)$ with the additional constraints (17), it is more efficient to index the moment matrices with a basis for $\mathbb{R}\langle\mathbf{x}\rangle_t$ modulo the ideal $\mathscr{I}_t(\{\sum_i v_i x_i : v \in \ker(A)\} \cup \{x_i x_j : A_{ij} = 0\})$.

## 2.4 Additional Properties of the Bounds

Here, we list some additional properties of the parameters $\xi_t^{\mathrm{cpsd}}(A)$ for $t \in \mathbb{N} \cup \{\infty, *\}$. First we state some properties for which the proofs are immediate and thus omitted.

**Lemma 4** *Suppose* $A \in \mathrm{CS}_+^n$ *and* $t \in \mathbb{N} \cup \{\infty, *\}$.

(1) *If* $P$ *is a permutation matrix, then* $\xi_t^{\mathrm{cpsd}}(A) = \xi_t^{\mathrm{cpsd}}(P^T A P)$.

(2) *If $B$ is a principal submatrix of $A$, then $\xi_t^{\text{cpsd}}(B) \leq \xi_t^{\text{cpsd}}(A)$.*

(3) *If $D$ is a positive definite diagonal matrix, then $\xi_t^{\text{cpsd}}(A) = \xi_t^{\text{cpsd}}(DAD)$.*

We also have the following direct sum property, where the equality follows using the $C^*$-algebra reformulations as given in Propositions 1 and 2.

**Lemma 5** *If $A \in \text{CS}_+^n$ and $B \in \text{CS}_+^m$, then $\xi_t^{\text{cpsd}}(A \oplus B) \leq \xi_t^{\text{cpsd}}(A) + \xi_t^{\text{cpsd}}(B)$, where equality holds for $t \in \{\infty, *\}$.*

**Proof** To prove the inequality, we take $L_A$ and $L_B$ feasible for $\xi_t^{\text{cpsd}}(A)$ and $\xi_t^{\text{cpsd}}(B)$, and construct a feasible $L$ for $\xi_t^{\text{cpsd}}(A \oplus B)$ by $L(p(\mathbf{x}, \mathbf{y})) = L_A(p(\mathbf{x}, \mathbf{0})) + L_B(p(\mathbf{0}, \mathbf{y}))$.

Now we show equality for $t = \infty$ ($t = *$). By Proposition 1 (Proposition 2), $\xi_t^{\text{cpsd}}(A \oplus B)$ is equal to the infimum over all $\alpha \geq 0$ for which there exists a (finite dimensional) unital $C^*$-algebra $\mathscr{A}$ with tracial state $\tau$ and $(\mathbf{X}, \mathbf{Y}) \in \mathscr{D}_{\mathscr{A}}(S_{A \oplus B}^{\text{cpsd}})$ such that $A = \alpha \cdot (\tau(X_i X_j))$, $B = \alpha \cdot (\tau(Y_i Y_j))$ and $(\tau(X_i Y_j)) = 0$. This implies $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S_A^{\text{cpsd}})$ and $\mathbf{Y} \in \mathscr{D}_{\mathscr{A}}(S_B^{\text{cpsd}})$. Let $P_A$ be the projection onto the space $\sum_i \text{Im}(X_i)$ and define the linear form $L_A \in \mathbb{R}\langle \mathbf{x} \rangle^*$ by $L_A(p) = \alpha \cdot \tau(p(\mathbf{X}) P_A)$. It follows that $L_A$ is nonnegative on $\mathscr{M}(S_A^{\text{cpsd}})$, and

$$L_A(x_i x_j) = \alpha \, \tau(x_i x_j P_A) = \alpha \, \tau(x_i x_j) = A_{ij},$$

so $L_A$ is feasible for $\xi_\infty^{\text{cpsd}}(A)$ with $L_A(1) = \alpha \tau(P_A)$. In the same way, we consider the projection $P_B$ onto the space $\sum_j \text{Im}(Y_j)$ and define a feasible solution $L_B$ for $\xi_t^{\text{cpsd}}(B)$ with $L_B(1) = \alpha \tau(P_B)$. By Lemma 12 we may assume $\tau$ to be faithful, so that positivity of $X_i$ and $Y_j$ together with $\tau(X_i Y_j) = 0$ implies $X_i Y_j = 0$ for all $i$ and $j$, and thus $\sum_i \text{Im}(X_i) \perp \sum_j \text{Im}(Y_j)$. This implies $I \succeq P_A + P_B$ and thus $\tau(P_A + P_B) \leq \tau(1) = 1$. We have

$$L_A(1) + L_B(1) = \alpha \, \tau(P_A) + \alpha \tau(P_B) \leq \alpha \, \tau(1) = \alpha,$$

so $\xi_t^{\text{cpsd}}(A) + \xi_t^{\text{cpsd}}(B) \leq L_A(1) + L_B(1) \leq \alpha$, completing the proof.  $\square$

Note that the cpsd-rank of a matrix satisfies the same properties as those mentioned in the above two lemmas, where the inequality in Lemma 5 is always an equality: cpsd-rank$_{\mathbb{C}}(A \oplus B) = $ cpsd-rank$_{\mathbb{C}}(A) + $ cpsd-rank$_{\mathbb{C}}(B)$ [38,62].

The following lemma shows that the first level of our hierarchy is at least as good as the analytic lower bound (18) on the cpsd-rank derived in [62, Theorem 10].

**Lemma 6** *For any non-zero matrix $A \in \text{CS}_+^n$, we have*

$$\xi_1^{\text{cpsd}}(A) \geq \frac{\left(\sum_{i=1}^n \sqrt{A_{ii}}\right)^2}{\sum_{i,j=1}^n A_{ij}}. \tag{18}$$

**Proof** Let $L$ be feasible for $\xi_1^{\mathrm{cpsd}}(A)$. Since $L$ is nonnegative on $\mathcal{M}_2(S_A^{\mathrm{cpsd}})$, it follows that $L(\sqrt{A_{ii}}x_i - x_i^2) \geq 0$, implying $\sqrt{A_{ii}}L(x_i) \geq L(x_i^2) = A_{ii}$ and thus $L(x_i) \geq \sqrt{A_{ii}}$. Moreover, the matrix $M_1(L)$ is positive semidefinite. By taking the Schur complement with respect to its upper left corner (indexed by 1), it follows that the matrix $L(1) \cdot A - (L(x_i)L(x_j))$ is positive semidefinite. Hence, the sum of its entries is nonnegative, which gives $L(1)(\sum_{i,j} A_{ij}) \geq (\sum_i L(x_i))^2 \geq (\sum_i \sqrt{A_{ii}})^2$ and shows the desired inequality.          □

As an application of Lemma 6, the first bound $\xi_1^{\mathrm{cpsd}}$ is exact for the $k \times k$ identity matrix: $\xi_1^{\mathrm{cpsd}}(I_k) = \mathrm{cpsd\text{-}rank}_{\mathbb{C}}(I_k) = k$. Moreover, by combining this with Lemma 4, it follows that $\xi_1^{\mathrm{cpsd}}(A) \geq k$ if $A$ contains a diagonal positive definite $k \times k$ principal submatrix. A slightly more involved example is given by the $5 \times 5$ circulant matrix $A$ whose entries are given by $A_{ij} = \cos((i-j)4\pi/5)^2$ ($i, j \in [5]$); this matrix was used in [25] to show a separation between the completely positive semidefinite cone and the completely positive cone, and it was shown that $\mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A) = 2$. The analytic lower bound of [62] also evaluates to 2, hence Lemma 6 shows that our bound is tight on this example.

We now examine further analytic properties of the parameters $\xi_t^{\mathrm{cpsd}}(\cdot)$. For each $r \in \mathbb{N}$, the set of matrices $A \in \mathrm{CS}_+^n$ with $\mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A) \leq r$ is closed, which shows that the function $A \mapsto \mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A)$ is lower semicontinuous. We now show that the functions $A \mapsto \xi_t^{\mathrm{cpsd}}(A)$ have the same property. The other bounds defined in this paper are also lower semicontinuous, with a similar proof.

**Lemma 7** *For every $t \in \mathbb{N} \cup \{\infty\}$ and $V \subseteq \mathbb{R}^n$, the function*

$$\mathrm{S}^n \to \mathbb{R} \cup \{\infty\}, \ A \mapsto \xi_{t,V}^{\mathrm{cpsd}}(A)$$

*is lower semicontinuous.*

**Proof** It suffices to show the result for $t \in \mathbb{N}$, because $\xi_{\infty,V}^{\mathrm{cpsd}}(A) = \sup_t \xi_{t,V}^{\mathrm{cpsd}}(A)$, and the pointwise supremum of lower semicontinuous functions is lower semicontinuous. We show that the level sets $\{A \in \mathrm{S}^n : \xi_{t,V}^{\mathrm{cpsd}}(A) \leq r\}$ are closed. For this, we consider a sequence $(A_k)_{k \in \mathbb{N}}$ in $\mathrm{S}^n$ converging to $A \in \mathrm{S}^n$ such that $\xi_{t,V}^{\mathrm{cpsd}}(A_k) \leq r$ for all $k$. We show that $\xi_{t,V}^{\mathrm{cpsd}}(A) \leq r$. Let $L_k \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ be an optimal solution to $\xi_{t,V}^{\mathrm{cpsd}}(A_k)$. As $L_k(1) \leq r$ for all $k$, it follows from Lemma 13 that there is a pointwise converging subsequence of $(L_k)_k$, still denoted $(L_k)_k$ for simplicity, that has a limit $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ with $L(1) \leq r$. To complete the proof we show that $L$ is feasible for $\xi_{t,V}^{\mathrm{cpsd}}(A)$. By the pointwise convergence of $L_k$ to $L$, for every $\varepsilon > 0$, $p \in \mathbb{R}\langle \mathbf{x} \rangle$, and $i \in [n]$, there exists a $K \in \mathbb{N}$ such that for all $k \geq K$ we have

$$|L(p^*x_i p) - L_k(p^*x_i p)| < \min\left\{1, \frac{\varepsilon}{\sqrt{A_{ii}}}\right\}, \qquad |L(p^*x_i^2 p) - L_k(p^*x_i^2 p)| < \varepsilon,$$

$$|\sqrt{A_{ii}} - \sqrt{(A_k)_{ii}}| < \frac{\varepsilon}{L(p^*x_i p) + 1}.$$

Hence, we have

$$L(p^*(\sqrt{A_{ii}}x_i - x_i^2)p) = \sqrt{A_{ii}}\Big(L(p^*x_i\,p) - L_k(p^*x_i\,p) + L_k(p^*x_i\,p)\Big)$$

$$- \Big(L(p^*x_i^2\,p) - L_k(p^*x_i^2\,p) + L_k(p^*x_i^2\,p)\Big)$$

$$\geq -2\varepsilon + \sqrt{A_{ii}}\,L_k(p^*x_i\,p) - L_k(p^*x_i^2\,p)$$

$$\geq -3\varepsilon + \sqrt{(A_k)_{ii}}\,L_k(p^*x_i\,p) - L_k(p^*x_i^2\,p)$$

$$= -3\varepsilon + L_k(p^*(\sqrt{(A_k)_{ii}}\,x_i - x_i^2)p) \geq -3\varepsilon,$$

where in the second inequality we use that $0 \leq L_k(p^*x_i\,p) \leq L(p^*x_i\,p) + 1$. Letting $\varepsilon \to 0$ gives $L(p^*(\sqrt{A_{ii}}x_i - x_i^2)p) \geq 0$.

Similarly, one can show $L(p^*(v^\mathsf{T}Av - (\sum_i v_i x_i)^2)p) \geq 0$ for $v \in V$, $p \in \mathbb{R}\langle \mathbf{x}\rangle$. ☐

If we restrict to completely positive semidefinite matrices with an all-ones diagonal, that is, to $\mathrm{CS}_+^n \cap \mathrm{E}_n$, we can show an even stronger property. Here, $\mathrm{E}_n$ is the *elliptope*, which is the set of $n \times n$ positive semidefinite matrices with an all-ones diagonal.

**Lemma 8** *For every $t \in \mathbb{N} \cup \{\infty\}$, the function*

$$\mathrm{CS}_+^n \cap \mathrm{E}_n \to \mathbb{R}, \ A \mapsto \xi_t^{\mathrm{cpsd}}(A)$$

*is convex, and hence continuous on the interior of its domain.*

**Proof** Let $A, B \in \mathrm{CS}_+^n \cap \mathrm{E}_n$ and $0 < \lambda < 1$. Let $L_A$ and $L_B$ be optimal solutions for $\xi_t^{\mathrm{cpsd}}(A)$ and $\xi_t^{\mathrm{cpsd}}(B)$. Since the diagonals of $A$ and $B$ are the same, we have $S_A^{\mathrm{cpsd}} = S_B^{\mathrm{cpsd}}$. So the linear functional $L = \lambda L_A + (1-\lambda)L_B$ is feasible for $\xi_t^{\mathrm{cpsd}}(\lambda A + (1-\lambda)B)$, hence $\xi_t^{\mathrm{cpsd}}(\lambda A + (1-\lambda)B) \leq \lambda L_A(1) + (1-\lambda)L_B(1) = \lambda\xi_t^{\mathrm{cpsd}}(A) + (1-\lambda)\xi_t^{\mathrm{cpsd}}(B)$. ☐

**Example 3** In this example, we show that for $t \geq 1$, the function

$$\mathrm{CS}_+^n \to \mathbb{R}, \ A \mapsto \xi_t^{\mathrm{cpsd}}(A)$$

is not continuous. For this, we consider the matrices

$$A_k = \begin{pmatrix} 1/k & 0 \\ 0 & 1 \end{pmatrix} \in \mathrm{CS}_+^2,$$

with $\mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A_k) = 2$ for all $k \geq 1$. As $A_k$ is diagonal positive definite, we have $\xi_t^{\mathrm{cpsd}}(A_k) = 2$ for all $t, k \geq 1$, while $\xi_t^{\mathrm{cpsd}}(\lim_{k\to\infty} A_k) = 1$. This argument extends to $\mathrm{CS}_+^n$ with $n > 2$. This example also shows that the first level of the hierarchy $\xi_1^{\mathrm{cpsd}}(\cdot)$ can be strictly better than the analytic lower bound (18) of [62].

**Example 4** In this example, we determine $\xi_t^{\mathrm{cpsd}}(A)$ for all $t \geq 1$ and $A \in \mathrm{CS}_+^2$. In view of Lemma 4(3), we only need to find $\xi_t^{\mathrm{cpsd}}(A(\alpha))$ for $0 \leq \alpha \leq 1$, where $A(\alpha) = \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}$.

The first bound $\xi_1^{\text{cpsd}}(A(\alpha))$ is equal to the analytic bound $2/(\alpha + 1)$ from (18), where the equality follows from the fact that $L$ given by $L(x_i x_j) = A(\alpha)_{ij}$, $L(x_1) = L(x_2) = 1$ and $L(1) = 2/(\alpha + 1)$ is feasible for $\xi_1^{\text{cpsd}}(A(\alpha))$.

For $t \geq 2$, we show $\xi_t^{\text{cpsd}}(A(\alpha)) = 2 - \alpha$. By the above, this is true for $\alpha = 0$ and $\alpha = 1$, and in Example 1 we show $\xi_t^{\text{cpsd}}(A(1/2)) = 3/2$ for $t \geq 2$. The claim then follows since the function $\alpha \mapsto \xi_t^{\text{cpsd}}(A(\alpha))$ is convex by Lemma 8.

## 3 Lower Bounds on the Completely Positive Rank

The best current approach for lower bounding the completely positive rank of a matrix is due to Fawzi and Parrilo [27]. Their approach relies on the atomicity of the completely positive rank, that is, the fact that cp-rank$(A) = r$ if and only if $A$ has an atomic decomposition $A = \sum_{k=1}^{r} v_k v_k^{\mathsf{T}}$ for nonnegative vectors $v_k$. In other words, if cp-rank$(A) = r$, then $A/r$ can be written as a convex combination of $r$ rank one positive semidefinite matrices $v_k v_k^{\mathsf{T}}$ that satisfy $0 \leq v_k v_k^{\mathsf{T}} \leq A$ and $v_k v_k^{\mathsf{T}} \preceq A$. Based on this observation, Fawzi and Parrilo define the parameter

$$\tau_{\text{cp}}(A) = \min\left\{\alpha : \alpha \geq 0, \ A \in \alpha \cdot \text{conv}\left\{R \in S^n : 0 \leq R \leq A, \ R \preceq A, \ \text{rank}(R) \leq 1\right\}\right\},$$

as lower bound for cp-rank$(A)$. They also define the semidefinite programming parameter

$$\tau_{\text{cp}}^{\text{sos}}(A) = \min\big\{\alpha : \alpha \in \mathbb{R}, \ X \in S^{n^2},$$
$$\begin{pmatrix} \alpha & \text{vec}(A)^{\mathsf{T}} \\ \text{vec}(A) & X \end{pmatrix} \succeq 0,$$
$$X_{(i,j),(i,j)} \leq A_{ij}^2 \quad \text{for} \quad 1 \leq i, j \leq n,$$
$$X_{(i,j),(k,l)} = X_{(i,l),(k,j)} \quad \text{for} \quad 1 \leq i < k \leq n, \ 1 \leq j < l \leq n,$$
$$X \preceq A \otimes A\big\},$$

as an efficiently computable relaxation of $\tau_{\text{cp}}(A)$, and they show rank$(A) \leq \tau_{\text{cp}}^{\text{sos}}(A)$. Therefore, we have

$$\text{rank}(A) \leq \tau_{\text{cp}}^{\text{sos}}(A) \leq \tau_{\text{cp}}(A) \leq \text{cp-rank}(A).$$

Instead of the atomic point of view, here we take the matrix factorization perspective, which allows us to obtain bounds by adapting the techniques from Sect. 2 to the commutative setting. Indeed, we may view a factorization $A = (a_i^T a_j)$ by nonnegative vectors as a factorization by diagonal (and thus pairwise commuting) positive semidefinite matrices.

Before presenting the details of our hierarchy of lower bounds, we mention some of our results in order to make the link to the parameters $\tau_{\text{cp}}^{\text{sos}}(A)$ and $\tau_{\text{cp}}(A)$. The direct analog of $\{\xi_t^{\text{cpsd}}(A)\}$ in the commutative setting leads to a hierarchy that does not

converge to $\tau_{cp}(A)$, but we provide two approaches to strengthen it that do converge to $\tau_{cp}(A)$. The first approach is based on a generalization of the tensor constraints in $\tau_{cp}^{sos}(A)$. We also provide a computationally more efficient version of these tensor constraints, leading to a hierarchy whose second level is at least as good as $\tau_{cp}^{sos}(A)$ while being defined by a smaller semidefinite program. The second approach relies on adding localizing constraints for vectors in the unit sphere as in Sect. 2.2.

The following hierarchy is a commutative analog of the hierarchy from Sect. 2, where we may now add the localizing polynomials $A_{ij} - x_i x_j$ for the pairs $1 \leq i < j \leq n$, which was not possible in the noncommutative setting of the completely positive semidefinite rank. For each $t \in \mathbb{N} \cup \{\infty\}$, we consider the semidefinite program

$$\xi_t^{cp}(A) = \min\{L(1) : L \in \mathbb{R}[x_1, \ldots, x_n]_{2t}^*,$$
$$L(x_i x_j) = A_{ij} \quad \text{for} \quad i, j \in [n],$$
$$L \geq 0 \quad \text{on} \quad \mathscr{M}_{2t}(S_A^{cp})\},$$

where we set

$$S_A^{cp} = \left\{\sqrt{A_{ii}} x_i - x_i^2 : i \in [n]\right\} \cup \left\{A_{ij} - x_i x_j : 1 \leq i < j \leq n\right\}.$$

We additionally define $\xi_*^{cp}(A)$ by adding the constraint $\text{rank}(M(L)) < \infty$ to $\xi_\infty^{cp}(A)$. We also consider the strengthening $\xi_{t,\dagger}^{cp}(A)$, where we add to $\xi_t^{cp}(A)$ the positivity constraints

$$L(gu) \geq 0 \quad \text{for} \quad g \in \{1\} \cup S_A^{cp} \quad \text{and} \quad u \in [\mathbf{x}]_{2t-\deg(g)} \tag{19}$$

and the tensor constraints

$$(L((ww')^c))_{w,w' \in \langle \mathbf{x} \rangle_{=l}} \preceq A^{\otimes l} \quad \text{for all integers} \quad 2 \leq l \leq t, \tag{20}$$

which generalize the case $l = 2$ used in the relaxation $\tau_{cp}^{sos}(A)$. Here, for a word $w \in \langle \mathbf{x} \rangle$, we denote by $w^c$ the corresponding (commutative) monomial in $[\mathbf{x}]$. The tensor constraints (20) involve matrices indexed by the *noncommutative* words of length exactly $l$. In Sect. 3.4, we show a more economical way to rewrite these constraints as $(L(mm'))_{m,m' \in [\mathbf{x}]_{=l}} \preceq Q_l A^{\otimes l} Q_l^\mathsf{T}$, thus involving smaller matrices indexed by *commutative* words of degree $l$.

Note that, as before, we can strengthen the bounds by adding other localizing polynomials to the set $S_A^{cp}$. In particular, we can follow the approach of Sect. 2.2. Another possibility is to add localizing constraints specific to the commutative setting: we can add each monomial $u \in [\mathbf{x}]$ to $S_A^{cp}$ (see Sect. 3.5.2 for an example).

The bounds $\xi_t^{cp}(A)$ and $\xi_{t,\dagger}^{cp}(A)$ are monotonically nondecreasing in $t$, and they are invariant under simultaneously permuting the rows and columns of $A$ and under scaling a row and column of $A$ by a positive number. In Propositions 6 and 7, we show

$$\tau_{cp}^{sos}(A) \leq \xi_{t,\dagger}^{cp}(A) \leq \tau_{cp}(A) \quad \text{for} \quad t \geq 2,$$

and in Proposition 10, we show the equality $\xi_{*,\dagger}^{cp}(A) = \tau_{cp}(A)$.

### 3.1 Comparison to $\tau_{cp}^{sos}(A)$

We first show that the semidefinite programs defining $\xi_{t,\dagger}^{cp}(A)$ are valid relaxations for the completely positive rank. More precisely, we show that they lower bound $\tau_{cp}(A)$.

**Proposition 6** *For $A \in CP^n$ and $t \in \mathbb{N} \cup \{\infty, *\}$, we have $\xi_{t,\dagger}^{cp}(A) \leq \tau_{cp}(A)$.*

**Proof** It suffices to show the inequality for $t = *$. For this, consider a decomposition $A = \alpha \sum_{k=1}^{r} \lambda_k R_k$, where $\alpha \geq 1$, $\lambda_k > 0$, $\sum_{k=1}^{r} \lambda_k = 1$, $0 \leq R_k \leq A$, $R_k \preceq A$, and rank $R_k = 1$. There are nonnegative vectors $v_k$ such that $R_k = v_k v_k^\mathsf{T}$. Define the linear map $L \in \mathbb{R}[\mathbf{x}]^*$ by $L = \alpha \sum_{k=1}^{r} \lambda_k L_{v_k}$, where $L_{v_k}$ is the evaluation at $v_k$ mapping any polynomial $p \in \mathbb{R}[\mathbf{x}]$ to $p(v_k)$.

The equality $(L(x_i x_j)) = A$ follows from the identity $A = \alpha \sum_{k=1}^{r} \lambda_k R_k$. The constraints $L((\sqrt{A_{ii}} x_i - x_i^2) p^2) \geq 0$ follow because

$$L_{v_k}(\sqrt{A_{ii}} x_i - x_i^2) p^2) = (\sqrt{A_{ii}}(v_k)_i - (v_k)_i^2) p(v_k)^2 \geq 0,$$

where we use that $(v_k)_i \geq 0$ and $(v_k)_i^2 = (R_k)_{ii} \leq A_{ii}$ implies $(v_k)_i^2 \leq (v_k)_i \sqrt{A_{ii}}$. The constraints $L((A_{ij} - x_i x_j) p^2) \geq 0$ and

$$L(gu) \geq 0 \quad \text{for} \quad g \in \{1\} \cup S_A^{cp} \quad \text{and} \quad u \in [\mathbf{x}]$$

follow in a similar way.

It remains to be shown that $X_l \preceq A^{\otimes l}$ for all $l$, where we set $X_l = (L(uv))_{u,v \in \langle \mathbf{x} \rangle_{=l}}$. Note that $X_1 = A$. We adapt the argument used in [27] to show $X_l \preceq A^{\otimes l}$ using induction on $l \geq 2$. Suppose $A^{\otimes(l-1)} \succeq X_{l-1}$. Combining $A - R_k \succeq 0$ and $R_k \succeq 0$ gives $(A - R_k) \otimes R_k^{\otimes(l-1)} \succeq 0$ and thus $A \otimes R_k^{\otimes(l-1)} \succeq R_k^{\otimes l}$ for each $k$. Scale by factor $\alpha \lambda_k$ and sum over $k$ to get

$$A \otimes X_{l-1} = \sum_k \alpha \lambda_k A \otimes R_k^{\otimes(l-1)} \succeq \sum_k \alpha \lambda_k R_k^{\otimes l} = X_l.$$

Finally, combining with $A^{\otimes(l-1)} - X_{l-1} \succeq 0$ and $A \succeq 0$, we obtain

$$A^{\otimes l} = A \otimes (A^{\otimes(l-1)} - X_{l-1}) + A \otimes X_{l-1} \succeq A \otimes X_{l-1} \succeq X_l.$$

$\square$

Now we show that the new parameter $\xi_{2,\dagger}^{cp}(A)$ is at least as good as $\tau_{cp}^{sos}(A)$. Later in Sect. 3.5.1, we will give an example where the inequality is strict.

**Proposition 7** *For $A \in CP^n$ we have $\tau_{cp}^{sos}(A) \leq \xi_{2,\dagger}^{cp}(A)$.*

**Proof** Let $L$ be feasible for $\xi_{2,\dagger}^{cp}(A)$. We will construct a feasible solution to the program defining $\tau_{cp}^{sos}(A)$ with objective value $L(1)$, which implies $\tau_{cp}^{sos}(A) \leq L(1)$ and thus

the desired inequality. For this set $\alpha = L(1)$ and define the symmetric $n^2 \times n^2$ matrix $X$ by $X_{(i,j),(k,l)} = L(x_i x_j x_k x_l)$ for $i, j, k, l \in [n]$. Then, the matrix

$$M := \begin{pmatrix} \alpha & \text{vec}(A)^{\mathsf{T}} \\ \text{vec}(A) & X \end{pmatrix}$$

is positive semidefinite. This follows because $M$ is obtained from the principal submatrix of $M_2(L)$ indexed by the monomials $1$ and $x_i x_j$ ($1 \le i \le j \le n$) where the rows/columns indexed by $x_j x_i$ with $1 \le i < j \le n$ are duplicates of the rows/columns indexed by $x_i x_j$.

We have $L((A_{ij} - x_i x_j) x_i x_j) \ge 0$ for all $i, j$: For $i \ne j$ this follows using the constraint $L((A_{ij} - x_i x_j) u) \ge 0$ with $u = x_i x_j$ (from (19)), and for $i = j$ this follows from

$$L((A_{ii} - x_i^2) x_i^2) = L\left( (\sqrt{A_{ii}} - x_i)^2 + 2(\sqrt{A_{ii}} x_i - x_i^2) \right) \ge 0,$$

which holds because of (10), the constraint $L(p^2) \ge 0$ for $\deg(p) \le 2$, and the constraint $L(\sqrt{A_{ii}} x_i - x_i^2) \ge 0$. Using $L(x_i x_j) = A_{ij}$, we get $X_{(i,j),(i,j)} = L(x_i^2 x_j^2) \le A_{ij}^2$. We also have $X_{(i,j),(k,l)} = L(x_i x_j x_k x_l) = L(x_i x_l x_k x_j) = X_{(i,l),(k,j)}$, and the constraint $(L(uv))_{u,v \in \langle \mathbf{x} \rangle_{=2}} \preceq A^{\otimes 2}$ implies $X \preceq A \otimes A$.  □

## 3.2 Convergence of the Basic Hierarchy

We first summarize convergence properties of the hierarchy $\xi_t^{\text{cp}}(A)$. Note that unlike in Sect. 2 where we can only claim the inequality $\xi_\infty^{\text{cpsd}}(A) \le \xi_*^{\text{cpsd}}(A)$, here we can show the equality $\xi_\infty^{\text{cp}}(A) = \xi_*^{\text{cp}}(A)$. This is because we can use Theorem 7, which permits to represent certain truncated linear functionals by finite atomic measures.

**Proposition 8** *Let $A \in CP^n$. For every $t \in \mathbb{N} \cup \{\infty, *\}$ the optimum in $\xi_t^{\text{cp}}(A)$ is attained, and $\xi_t^{\text{cp}}(A) \to \xi_\infty^{\text{cp}}(A) = \xi_*^{\text{cp}}(A)$ as $t \to \infty$. If $\xi_t^{\text{cp}}(A)$ admits a flat optimal solution, then $\xi_t^{\text{cp}}(A) = \xi_\infty^{\text{cp}}(A)$. Moreover, $\xi_\infty^{\text{cp}}(A) = \xi_*^{\text{cp}}(A)$ is the minimum value of $L(1)$ taken over all conic combinations $L$ of evaluations at elements of $D(S_A^{\text{cp}})$ satisfying $A = (L(x_i x_j))$.*

**Proof** We may assume $A \ne 0$. Since $\sqrt{A_{ii}} x_i - x_i^2 \in S_A^{\text{cp}}$ for all $i$, using (10) we obtain that $\text{Tr}(A) - \sum_i x_i^2 \in \mathcal{M}_2(S_A^{\text{cp}})$. By adapting the proof of Proposition 1 to the commutative setting, we see that the optimum in $\xi_t^{\text{cp}}(A)$ is attained for $t \in \mathbb{N} \cup \{\infty\}$, and $\xi_t^{\text{cp}}(A) \to \xi_\infty^{\text{cp}}(A)$ as $t \to \infty$.

We now show the inequality $\xi_*^{\text{cp}}(A) \le \xi_\infty^{\text{cp}}(A)$, which implies that equality holds. For this, let $L$ be optimal for $\xi_\infty^{\text{cp}}(A)$. By Theorem 7, the restriction of $L$ to $\mathbb{R}[\mathbf{x}]_2$ extends to a conic combination of evaluations at points in $D(S_A^{\text{cp}})$. It follows that this extension is feasible for $\xi_*^{\text{cp}}(A)$ with the same objective value. This shows that $\xi_*^{\text{cp}}(A) \le \xi_\infty^{\text{cp}}(A)$, that the optimum in $\xi_*^{\text{cp}}(A)$ is attained, and that $\xi_*^{\text{cp}}(A)$ is the minimum of $L(1)$ over all conic combinations $L$ of evaluations at elements of $D(S_A^{\text{cp}})$ such that $A = (L(x_i x_j))$. Finally, by Theorem 6, we have $\xi_t^{\text{cp}}(A) = \xi_\infty^{\text{cp}}(A)$ if $\xi_t^{\text{cp}}(A)$ admits a flat optimal solution.  □

Next, we give a reformulation for the parameter $\xi_*^{cp}(A)$, which is similar to the formulation of $\tau_{cp}(A)$, although it lacks the constraint $R \preceq A$ which is present in $\tau_{cp}(A)$.

**Proposition 9** *We have*

$$\xi_*^{cp}(A) = \min\left\{\alpha : \alpha \geq 0, \ A \in \alpha \cdot \text{conv}\left\{R \in S^n : 0 \leq R \leq A, \ \text{rank}(R) \leq 1\right\}\right\}.$$

**Proof** This follows directly from the reformulation of $\xi_*^{cp}(A)$ in Proposition 8 in terms of conic evaluations at points in $D(S_A^{cp})$ after observing that, for $v \in \mathbb{R}^n$, we have $v \in D(S_A^{cp})$ if and only if the matrix $R = vv^\top$ satisfies $0 \leq R \leq A$. $\qquad\square$

### 3.3 Additional Constraints and Convergence to $\tau_{cp}(A)$

The reformulation of the parameter $\xi_*^{cp}(A)$ in Proposition 9 differs from $\tau_{cp}(A)$ in that the constraint $R \preceq A$ is missing. In order to have a hierarchy converging to $\tau_{cp}(A)$, we need to add constraints to enforce that $L$ can be decomposed as a conic combination of evaluation maps at nonnegative vectors $v$ satisfying $vv^T \preceq A$. Here, we present two ways to achieve this goal. First, we show that the tensor constraints (20) suffice in the sense that $\xi_{*,\dagger}^{cp}(A) = \tau_{cp}(A)$ (note that the constraints (19) are not needed for this result). However, because of the special form of the tensor constraints we do not know whether $\xi_{t,\dagger}^{cp}(A)$ admitting a flat optimal solution implies $\xi_{t,\dagger}^{cp}(A) = \xi_{*,\dagger}^{cp}(A)$, and we do not know whether $\xi_{\infty,\dagger}^{cp}(A) = \xi_{*,\dagger}^{cp}(A)$. Second, we adapt the approach of adding additional localizing constraints from Sect. 2.2 to the commutative setting, where we do show $\xi_{\infty,\mathbb{S}^{n-1}}^{cp}(A) = \xi_{*,\mathbb{S}^{n-1}}^{cp}(A) = \tau_{cp}(A)$. This yields a doubly indexed sequence of semidefinite programs whose optimal values converge to $\tau_{cp}(A)$.

**Proposition 10** *Let $A \in CP^n$. For every $t \in \mathbb{N} \cup \{\infty\}$, the optimum in $\xi_{t,\dagger}^{cp}(A)$ is attained. We have $\xi_{t,\dagger}^{cp}(A) \to \xi_{\infty,\dagger}^{cp}(A)$ as $t \to \infty$ and $\xi_{*,\dagger}^{cp}(A) = \tau_{cp}(A)$.*

**Proof** The attainment of the optima in $\xi_{t,\dagger}^{cp}(A)$ for $t \in \mathbb{N} \cup \{\infty\}$ and the convergence of $\xi_{t,\dagger}^{cp}(A)$ to $\xi_{\infty,\dagger}^{cp}(A)$ can be shown in the same way as the analog statements for $\xi_t^{cp}(A)$ in Proposition 8.

We have seen the inequality $\xi_{*,\dagger}^{cp}(A) \leq \tau_{cp}(A)$ in Proposition 6. Now we show the reverse inequality. Let $L$ be feasible for $\xi_{*,\dagger}^{cp}(A)$. We will show that $L$ is feasible for $\tau_{cp}(A)$, which implies $\tau_{cp}(A) \leq L(1)$ and thus $\tau_{cp}(A) \leq \xi_{*,\dagger}^{cp}(A)$.

By Proposition 7 and the fact that $\text{rank}(A) \leq \tau_{cp}^{sos}(A)$, we have $L(1) > 0$ (where we assume $A \neq 0$). By Theorem 5, we may write

$$L = L(1) \sum_{k=1}^{K} \lambda_k L_{v_k},$$

where $\lambda_k > 0$, $\sum_k \lambda_k = 1$, and $L_{v_k}$ is an evaluation map at a point $v_k \in D(S_A^{cp})$. We define the matrices $R_k = v_k v_k^\top$, so that $A = L(1) \sum_{k=1}^{K} R_k$. The matrices $R_k$ satisfy $0 \leq R_k \leq A$ since $v_k \in D(S_A^{cp})$. Clearly also $R_k \succeq 0$. It remains to show that $R_k \preceq A$.

For this we use the tensor constraints (20). Using that $L$ is a conic combination of evaluation maps, we may rewrite these constraints as

$$L(1) \sum_{k=1}^{K} \lambda_k R_k^{\otimes l} \preceq A^{\otimes l},$$

from which it follows that $L(1)\lambda_k R_k^{\otimes l} \preceq A^{\otimes l}$ for all $k \in [K]$. Therefore, for all $k \in [K]$ and all vectors $v$ with $v^T R_k v > 0$, we have

$$L(1)\lambda_k \leq \left( \frac{v^T A v}{v^T R_k v} \right)^l \quad \text{for all} \quad l \in \mathbb{N}. \tag{21}$$

Suppose there is a $k$ such that $R_k \npreceq A$. Then there exists a $v$ such that $v^T R_k v > v^T A v$. As $(v^T A v)/(v^T R_k v) < 1$, letting $l$ tend to $\infty$ we obtain $L(1)\lambda_k = 0$, reaching a contradiction. It follows that $R_k \preceq A$ for all $k \in [K]$. $\qquad\square$

The second approach for reaching $\tau_{cp}(A)$ is based on using the extra localizing constraints from Sect. 2.2. For a subset $V \subseteq \mathbb{S}^{n-1}$, define $\xi_{t,V}^{cp}(A)$ by replacing the truncated quadratic module $\mathcal{M}_{2t}(S_A^{cp})$ in $\xi_t^{cp}(A)$ by $\mathcal{M}_{2t}(S_{A,V}^{cp})$, where

$$S_{A,V}^{cp} = S_A^{cp} \cup \left\{ v^T A v - \left( \sum_{i=1}^{n} v_i x_i \right)^2 : v \in V \right\}.$$

Proposition 5 can be adapted to the completely positive setting, so that we have a sequence of finite subsets $V_1 \subseteq V_2 \subseteq \ldots \subseteq \mathbb{S}^{n-1}$ with $\xi_{*,V_k}^{cp}(A) \to \xi_{*,\mathbb{S}^{n-1}}^{cp}(A)$ as $k \to \infty$. Proposition 8 still holds when adding extra localizing constraints, so that for any $k \geq 1$ we have

$$\lim_{t \to \infty} \xi_{t,V_k}^{cp}(A) = \xi_{*,V_k}^{cp}(A).$$

Combined with Proposition 11, this shows that we have a doubly indexed sequence $\xi_{t,V_k}^{cp}(A)$ of semidefinite programs that converges to $\tau_{cp}(A)$ as $t \to \infty$ and $k \to \infty$.

**Proposition 11** *For $A \in CP^n$ we have $\xi_{*,\mathbb{S}^{n-1}}^{cp}(A) = \tau_{cp}(A)$.*

**Proof** The proof is the same as the proof of Proposition 9, with the following additional observation: Given a vector $u \in \mathbb{R}^n$, we have $u \in D(S_{A,\mathbb{S}^{n-1}}^{cp})$ only if $uu^T \preceq A$. The latter follows from the additional localizing constraints: for each $v \in \mathbb{R}^n$ we have

$$0 \leq v^T A v - \left( \sum_i v_i u_i \right)^2 = v^T (A - uu^T) v. \qquad\square$$

### 3.4 More Efficient Tensor Constraints

Here, we show that for any integer $l \geq 2$ the constraint $A^{\otimes l} - (L((ww')^c))_{w,w' \in \langle \mathbf{x} \rangle_{=l}} \succeq 0$, used in the definition of $\xi_{t,+}^{\mathrm{cp}}(A)$, can be reformulated in a more economical way using matrices indexed by *commutative* monomials in $[\mathbf{x}]_{=l}$ instead of noncommutative words in $\langle \mathbf{x} \rangle_{=l}$. For this we exploit the symmetry in the matrices $A^{\otimes l}$ and $(L((ww')^c))_{w,w' \in \langle \mathbf{x} \rangle_{=l}}$ for $L \in \mathbb{R}[\mathbf{x}]_{2l}^*$. Recall that for a word $w \in \langle \mathbf{x} \rangle$, we let $w^c$ denote the corresponding (commutative) monomial in $[\mathbf{x}]$.

Define the matrix $Q_l \in \mathbb{R}^{[\mathbf{x}]_{=l} \times \langle \mathbf{x} \rangle_{=l}}$ by

$$(Q_l)_{m,w} = \begin{cases} 1/d_m & \text{if } w^c = m, \\ 0 & \text{otherwise,} \end{cases} \tag{22}$$

where, for $m = x_1^{\alpha_1} \cdots x_n^{\alpha_n} \in [\mathbf{x}]_{=l}$, we define the multinomial coefficient

$$d_m = \left| \left\{ w \in \langle \mathbf{x} \rangle_{=l} : w^c = m \right\} \right| = \frac{l!}{\alpha_1! \cdots \alpha_n!}. \tag{23}$$

**Lemma 9** *For $L \in \mathbb{R}[\mathbf{x}]_{2l}^*$ we have*

$$Q_l (L((ww')^c))_{w,w' \in \langle \mathbf{x} \rangle_{=l}} Q_l^T = (L(mm'))_{m,m' \in [\mathbf{x}]_{=l}}.$$

**Proof** For $m, m' \in [\mathbf{x}]_l$, the $(m, m')$-entry of the left hand side is equal to

$$\sum_{w,w' \in \langle \mathbf{x} \rangle_{=l}} Q_{mw} Q_{m'w'} L((ww')^c) = \sum_{\substack{w \in \langle \mathbf{x} \rangle_{=l} \\ w^c = m}} \sum_{\substack{w' \in \langle \mathbf{x} \rangle_{=l} \\ (w')^c = m'}} \frac{L((ww')^c)}{d_m d_{m'}} = L(mm').$$

$\square$

The symmetric group $S_l$ acts on $\langle \mathbf{x} \rangle_{=l}$ by $(x_{i_1} \cdots x_{i_l})^\sigma = x_{i_{\sigma(1)}} \cdots x_{i_{\sigma(l)}}$ for $\sigma \in S_l$. Let

$$P = \frac{1}{l!} \sum_{\sigma \in S_l} P_\sigma, \tag{24}$$

where, for any $\sigma \in S_l$, $P_\sigma \in \mathbb{R}^{\langle \mathbf{x} \rangle_{=l} \times \langle \mathbf{x} \rangle_{=l}}$ is the permutation matrix defined by

$$(P_\sigma)_{w,w'} = \begin{cases} 1 & \text{if } w^\sigma = w', \\ 0 & \text{otherwise.} \end{cases}$$

A matrix $M \in \mathbb{R}^{\langle \mathbf{x} \rangle_{=l} \times \langle \mathbf{x} \rangle_{=l}}$ is said to be $S_l$-*invariant* if $P^\sigma M = M P^\sigma$ for all $\sigma \in S_l$.

**Lemma 10** *If $M \in \mathbb{R}^{\langle \mathbf{x} \rangle_{=l} \times \langle \mathbf{x} \rangle_{=l}}$ is symmetric and $S_l$-invariant, then*

$$M \succeq 0 \iff Q_l M Q_l^T \succeq 0.$$

**Proof** The implication $M \succeq 0 \implies Q_l M Q_l^\top \succeq 0$ is immediate. For the other implication, we need a preliminary fact. Consider the diagonal matrix $D \in \mathbb{R}^{[\mathbf{x}]_{=l} \times [\mathbf{x}]_{=l}}$ with $D_{mm} = d_m$ for $m \in [\mathbf{x}]_{=l}$. We claim that $Q_l^\top D Q_l = P$, the matrix in (24). Indeed, for any $w, w' \in \langle \mathbf{x} \rangle_{=l}$, we have

$$(Q_l^\top D Q_l)_{ww'} = \sum_{m \in [\mathbf{x}]_{=l}} (Q_l)_{mw} (Q_l)_{mw'} D_{mm} = \begin{cases} 1/d_m & \text{if } w^c = (w')^c = m, \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{|\{\sigma \in S_l : w^\sigma = w'\}|}{l!} = P_{ww'}.$$

Suppose $Q_l M Q_l^\top \succeq 0$, and let $\lambda$ be an eigenvalue of $M$ with eigenvector $z$. Since $MP = PM$, we may assume $Pz = z$, for otherwise we can replace $z$ by $Pz$, which is still an eigenvector of $M$ with eigenvalue $\lambda$. We may also assume $z$ to be a unit vector. Then, $\lambda \geq 0$ can be shown using the identity $Q_l^\top D Q_l = P$ as follows:

$$\lambda = z^\top M z = z^\top P M P z = z^\top (Q_l^\top D Q_l) M (Q_l^\top D Q_l) z = (DQ_l z)^\top (Q_l M Q_l^\top) DQ_l z \geq 0.$$

$\square$

We can now derive our symmetry reduction result:

**Proposition 12** *For $L \in \mathbb{R}[\mathbf{x}]_{2l}^*$ we have*

$$A^{\otimes l} - (L((ww')^c))_{w,w' \in \langle \mathbf{x} \rangle_{=l}} \succeq 0 \iff Q_l A^{\otimes l} Q_l^\top - (L(mm'))_{m,m' \in [\mathbf{x}]_{=l}} \succeq 0.$$

**Proof** For any $w, w' \in \langle \mathbf{x} \rangle_{=l}$, we have $(P_\sigma A^{\otimes l} P_\sigma^\top)_{w,w'} = A^{\otimes l}_{w^\sigma, (w')^\sigma} = A^{\otimes l}_{w,w'}$ and

$$(P_\sigma (L((uu')^c))_{u,u' \in \langle \mathbf{x} \rangle_{=l}} P_\sigma^*)_{w,w'} = L((w^\sigma (w')^\sigma)^c) = L((ww')^c).$$

This shows that the matrix $A^{\otimes l} - (L((ww')^c))_{w,w' \in \langle \mathbf{x} \rangle_{=l}}$ is $S_l$-invariant. Hence, the claimed result follows by using Lemmas 9 and 10. $\square$

### 3.5 Computational Examples

### 3.5.1 Bipartite Matrices

Consider the $(p + q) \times (p + q)$ matrices

$$P(a, b) = \begin{pmatrix} (a + q)I_p & J_{p,q} \\ J_{q,p} & (b + p)I_q \end{pmatrix}, \quad a, b \in \mathbb{R}_+,$$

where $J_{p,q}$ denotes the all-ones matrix of size $p \times q$. We have $P(a, b) = P(0, 0) + D$ for some nonnegative diagonal matrix $D$. As can be easily verified, $P(0, 0)$ is completely positive with cp-rank$(P(0, 0)) = pq$, so $P(a, b)$ is completely positive with $pq \leq$ cp-rank$(P(a, b)) \leq pq + p + q$.

For $p = 2$ and $q = 3$, we have cp-rank$(P(a, b)) = 6$ for all $a, b \geq 0$, which follows from the fact that $5 \times 5$ completely positive matrices with at least one zero entry have cp-rank at most 6; see [6, Theorem 3.12]. Fawzi and Parrilo [27] show that $\tau_{\mathrm{cp}}^{\mathrm{sos}}(P(0, 0)) = 6$, and give a subregion of $[0, 1]^2$ where $5 < \tau_{\mathrm{cp}}^{\mathrm{sos}}(P(a, b)) < 6$. The next lemma shows the bound $\xi_{2,\dagger}^{\mathrm{cp}}(P(a, b))$ is tight for all $a, b \geq 0$ and therefore strictly improves on $\tau_{\mathrm{cp}}^{\mathrm{sos}}$ in this region.

**Lemma 11** *For $a, b \geq 0$ we have $\xi_{2,\dagger}^{\mathrm{cp}}(P(a, b)) \geq pq$.*

**Proof** Let $L$ be feasible for $\xi_{2,\dagger}^{\mathrm{cp}}(P(a, b))$ and let

$$B = \begin{pmatrix} \alpha & c^\mathsf{T} \\ c & X \end{pmatrix}$$

be the principal submatrix of $M_2(L)$ where the rows and columns are indexed by

$$\{1\} \cup \{x_i x_j : 1 \leq i \leq p, \ p + 1 \leq j \in p + q\}.$$

It follows that $c$ is the all-ones vector $c = \mathbf{1}$. Moreover, if $P(a, b)_{ij} = 0$ for some $i \neq j$, then the constraints $L(x_i x_j u) \geq 0$ and $L((P(a, b)_{ij} - x_i x_j)u) \geq 0$ imply $L(x_i x_j u) = 0$ for all $u \in [\mathbf{x}]_2$. Hence, $X_{x_i x_j, x_k x_l} = L(x_i x_j x_k x_l) = 0$ whenever $x_i x_j \neq x_k x_l$. It follows that $X$ is a diagonal matrix. We write

$$B = \begin{pmatrix} \alpha & \mathbf{1}^\mathsf{T} \\ \mathbf{1} & \mathrm{Diag}(z_1, \ldots, z_{pq}) \end{pmatrix}.$$

Since $\begin{pmatrix} 1 & -\mathbf{1}^\mathsf{T} \\ -\mathbf{1} & J \end{pmatrix} \succeq 0$ we have

$$0 \leq \mathrm{Tr}\left( \begin{pmatrix} \alpha & \mathbf{1}^\mathsf{T} \\ \mathbf{1} & \mathrm{Diag}(z_1, \ldots, z_{pq}) \end{pmatrix} \begin{pmatrix} 1 & -\mathbf{1}^\mathsf{T} \\ -\mathbf{1} & J \end{pmatrix} \right) = \alpha - 2pq + \sum_{k=1}^{pq} z_k.$$

Finally, by the constraints $L((P(a, b)_{ij} - x_i x_j)u) \geq 0$ (with $i \in [p], j \in p + [q]$ and $u = x_i x_j$) and $L(x_i x_j) = P(a, b)_{ij}$ we obtain $z_k \leq 1$ for all $k \in [pq]$. Combined with the above inequality, it follows that

$$L(1) = \alpha \geq 2pq - \sum_{k=1}^{pq} z_k \geq pq,$$

and hence $\xi_{2,\dagger}^{\mathrm{cp}}(P(a, b)) \geq pq$. $\qquad\square$

**Table 1** Examples from [10] with various bounds on their cp-rank

| Example | cp-rank$(\cdot)$ | $\lfloor \frac{n^2}{4} \rfloor$ | rank$(\cdot)$ | $\xi_1^{cp}(\cdot)$ | $\xi_2^{cp}(\cdot)$ | $\xi_{2,\dagger}^{cp}(\cdot)$ | $\xi_{2,\dagger}^{cp}(\cdot) + x_i x_j$ | $\xi_{3,\dagger}^{cp}(\cdot)$ |
|---|---|---|---|---|---|---|---|---|
| $M_7$ | 14 | 12 | 7 | 2.64 | 4.21 | 7.21 | 9.75 | 10.50 |
| $\widetilde{M}_7$ | 14 | 12 | 7 | 2.58 | 4.66 | 8.43 | 9.53 | 10.50 |
| $\widetilde{M}_8$ | 18 | 16 | 8 | 3.23 | 5.45 | 8.74 | 10.41 | 13.82 |
| $\widetilde{M}_9$ | 26 | 20 | 9 | 3.39 | 5.71 | 11.60 | 13.74 | 17.74 |
| $M_{11}$ | 32 | 30 | 11 | 4.32 | 7.46 | 20.76 | 21.84 | – |

### 3.5.2 Examples Related to the DJL-Conjecture

The Drew–Johnson–Loewy conjecture [21] states that the maximal cp-rank of an $n \times n$ completely positive matrix is equal to $\lfloor n^2/4 \rfloor$. Recently, this conjecture has been disproven for $n = 7, 8, 9, 10, 11$ in [10] and for all $n \geq 12$ in [11] (interestingly, it remains open for $n = 6$). Here, we study our bounds on the examples of [10]. Although our bounds are not tight for the cp-rank, they are non-trivial and as such may be of interest for future comparisons. For numerical stability reasons, we have evaluated our bounds on scaled versions of the matrices from [10], so that the diagonal entries become equal to 1. In Table 1 the matrices $\widetilde{M}_7$, $\widetilde{M}_8$ and $\widetilde{M}_9$ correspond to the matrices $\widetilde{M}$ in Examples 1–3 of [10], and $M_7$, $M_{11}$ correspond to the matrices $M$ in Examples 1 and 4. The column $\xi_{2,\dagger}^{cp}(\cdot) + x_i x_j$ corresponds to the bound $\xi_{2,\dagger}^{cp}(\cdot)$ where we replace $S_A^{cp}$ by $S_A^{cp} \cup \{x_i x_j : 1 \leq i < j \leq n\}$.

## 4 Lower Bounds on the Nonnegative Rank

In this section, we adapt the techniques for the cp-rank from Sect. 3 to the asymmetric setting of the nonnegative rank. We now view a factorization $A = (a_i^\mathsf{T} b_j)_{i \in [m], j \in [n]}$ by nonnegative vectors as a factorization by positive semidefinite diagonal matrices, by writing $A_{ij} = \mathrm{Tr}(X_i X_{m+j})$, with $X_i = \mathrm{Diag}(a_i)$ and $X_{m+j} = \mathrm{Diag}(b_j)$. Note that we can view this as a "partial matrix" setting, where for the symmetric matrix $(\mathrm{Tr}(X_i X_k))_{i,k \in [m+n]}$ of size $m + n$, only the off-diagonal entries at the positions $(i, m + j)$ for $i \in [m]$, $j \in [n]$ are specified.

This asymmetry requires rescaling the factors in order to get upper bounds on their maximal eigenvalues, which is needed to ensure the Archimedean property for the selected localizing polynomials. For this we use the well-known fact that for any $A \in \mathbb{R}_+^{m \times n}$ there exists a factorization $A = (\mathrm{Tr}(X_i X_{m+j}))$ by diagonal nonnegative matrices of size rank$_+(A)$, such that

$$\lambda_{\max}(X_i), \lambda_{\max}(X_{m+j}) \leq \sqrt{A_{\max}} \quad \text{for all} \quad i \in [m], j \in [n],$$

where $A_{\max} := \max_{i,j} A_{ij}$. To see this, observe that for any rank one matrix $R = uv^\mathsf{T}$ with $0 \leq R \leq A$, one may assume $0 \leq u_i, v_j \leq \sqrt{A_{\max}}$ for all $i, j$. Hence, the set

$$S_A^+ = \left\{ \sqrt{A_{\max}} x_i - x_i^2 : i \in [m+n] \right\} \cup \left\{ A_{ij} - x_i x_{m+j} : i \in [m], j \in [n] \right\}$$

is localizing for $A$; that is, there exists a minimal factorization $\mathbf{X}$ of $A$ with $\mathbf{X} \in \mathscr{D}(S_A^+)$.

Given $A \in \mathbb{R}_{\geq 0}^{m \times n}$, for each $t \in \mathbb{N} \cup \{\infty\}$ we consider the semidefinite program

$$
\begin{aligned}
\xi_t^+(A) = \min\big\{ L(1) : \; & L \in \mathbb{R}[x_1, \ldots, x_{m+n}]_{2t}^*, \\
& L(x_i x_{m+j}) = A_{ij} \quad \text{for} \quad i \in [m], j \in [n], \\
& L \geq 0 \quad \text{on} \quad \mathscr{M}_{2t}(S_A^+) \big\}.
\end{aligned}
$$

Moreover, define $\xi_*^+(A)$ by adding the constraint $\operatorname{rank}(M(L)) < \infty$ to the program defining $\xi_\infty^+(A)$. It is easy to check that $\xi_t^+(A) \leq \xi_\infty^+(A) \leq \xi_*^+(A) \leq \operatorname{rank}_+(A)$ for $t \in \mathbb{N}$.

Denote by $\xi_{t,\uparrow}^+(A)$ the strengthening of $\xi_t^+(A)$ where we add the positivity constraints

$$L(gu) \geq 0 \quad \text{for} \quad g \in \{1\} \cup S_A^+ \quad \text{and} \quad u \in [\mathbf{x}]_{2t - \deg(g)}. \tag{25}$$

Note that these extra constraints can help for finite $t$, but are redundant for $t \in \{\infty, *\}$.

## 4.1 Comparison to Other Bounds

As in the previous section, we compare our bounds to the bounds by Fawzi and Parrilo [27]. They introduce the following parameter $\tau_+(A)$ as analog of the bound $\tau_{\mathrm{cp}}(A)$ for the nonnegative rank:

$$\tau_+(A) = \min\left\{ \alpha : \alpha \geq 0, \; A \in \alpha \cdot \operatorname{conv}\left\{ R \in \mathbb{R}^{m \times n} : 0 \leq R \leq A, \; \operatorname{rank}(R) \leq 1 \right\} \right\},$$

and the analog $\tau_+^{\mathrm{sos}}(A)$ of the bound $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$ for the nonnegative rank:

$$
\begin{aligned}
\tau_+^{\mathrm{sos}}(A) = \inf\Big\{ \alpha : \; & X \in \mathbb{R}^{mn \times mn}, \; \alpha \in \mathbb{R}, \\
& \begin{pmatrix} \alpha & \operatorname{vec}(A)^{\mathsf{T}} \\ \operatorname{vec}(A) & X \end{pmatrix} \succeq 0, \\
& X_{(i,j),(i,j)} \leq A_{ij}^2 \quad \text{for} \quad 1 \leq i \leq m, 1 \leq j \leq n, \\
& X_{(i,j),(k,l)} = X_{(i,l),(k,j)} \quad \text{for} \quad 1 \leq i < k \leq m, \; 1 \leq j < l \leq n \Big\}.
\end{aligned}
$$

First, we give the analog of Proposition 8, whose proof we omit since it is very similar.

**Proposition 13** *Let $A \in \mathbb{R}_+^{m \times n}$. For every $t \in \mathbb{N} \cup \{\infty, *\}$ the optimum in $\xi_t^+(A)$ is attained, and $\xi_t^+(A) \to \xi_\infty^+(A) = \xi_*^+(A)$ as $t \to \infty$. If $\xi_t^+(A)$ admits a flat optimal solution, then $\xi_t^+(A) = \xi_*^+(A)$. Moreover, $\xi_\infty^+(A) = \xi_*^+(A)$ is the minimum of $L(1)$ over all conic combinations $L$ of trace evaluations at elements of $D(S_A^+)$ satisfying $A = (L(x_i x_{m+j}))$.*

Now we observe that the parameters $\xi_\infty^+(A)$ and $\xi_*^+(A)$ coincide with $\tau_+(A)$, so that we have a sequence of semidefinite programs converging to $\tau_+(A)$.

**Proposition 14** *For any $A \in \mathbb{R}_{\geq 0}^{m \times n}$, we have $\xi_\infty^+(A) = \xi_*^+(A) = \tau_+(A)$.*

**Proof** The discussion at the beginning of Sect. 4 shows that for any rank one matrix $R$ satisfying $0 \leq R \leq A$ we may assume that $R = uv^\mathsf{T}$ with $(u, v) \in \mathbb{R}_+^m \times \mathbb{R}_+^n$ and $u_i, v_j \leq \sqrt{A_{\max}}$ for $i \in [m]$, $j \in [n]$. Hence, $\tau_+(A)$ can be written as:

$$\min\left\{\alpha : \alpha \geq 0,\ A \in \alpha \cdot \mathrm{conv}\left\{uv^\mathsf{T} : u \in \left[0, \sqrt{A_{\max}}\right]^m, v \in \left[0, \sqrt{A_{\max}}\right]^n, uv^\mathsf{T} \leq A\right\}\right\}$$
$$= \min\left\{\alpha : \alpha \geq 0,\ A \in \alpha \cdot \mathrm{conv}\left\{uv^\mathsf{T} : (u, v) \in D(S_A^+)\right\}\right\}.$$

The equality $\xi_\infty^+(A) = \xi_*^+(A) = \tau_+(A)$ now follows from the reformulation of $\xi_*^+(A)$ in Proposition 13 in terms of conic evaluations, after noting that for $(u, v)$ in $\mathbb{R}^m \times \mathbb{R}^n$ we have $(u, v) \in D(S_A^+)$ if and only if the matrix $R = uv^\mathsf{T}$ satisfies $0 \leq R \leq A$. □

Analogously to the case of the completely positive rank, we have the following proposition. The proof is similar to that of Proposition 4.2, considering now for $M$ the principal submatrix of $M_2(L)$ indexed by the monomials 1 and $x_i x_{m+j}$ for $i \in [m]$ and $j \in [n]$.

**Proposition 15** *If $A$ is a nonnegative matrix, then $\xi_{2,\dagger}^+(A) \geq \tau_+^{\mathrm{sos}}(A)$.*

In the remainder of this section, we recall how $\tau_+(A)$ and $\tau_+^{\mathrm{sos}}(A)$ compare to other bounds in the literature. These bounds can be divided into two categories: combinatorial lower bounds and norm-based lower bounds. The following diagram from [27] summarizes how $\tau_+^{\mathrm{sos}}(A)$ and $\tau_+(A)$ relate to the combinatorial lower bounds

$$
\begin{array}{ccccc}
\tau_+^{\mathrm{sos}}(A) & \leq & \tau_+(A) & & \leq \mathrm{rank}_+(A) \\
\text{VI} & & \text{VI} & & \text{VI}
\end{array}
$$
$$\mathrm{fool}(A) = \omega(\mathrm{RG}(A)) \leq \overline{\vartheta}(\mathrm{RG}(A)) \leq \chi_{\mathrm{frac}}(\mathrm{RG}(A)) \leq \chi(\mathrm{RG}(A)) = \mathrm{rank}_B(A).$$

Here $\mathrm{RG}(A)$ is the *rectangular graph*, with $V = \{(i, j) \in [m] \times [n] : A_{ij} > 0\}$ as vertex set and $E = \{((i, j), (k, l)) : A_{il} A_{kj} = 0\}$ as edge set. The coloring number of $\mathrm{RG}(A)$ coincides with the well-known *rectangle covering number* (also denoted $\mathrm{rank}_B(A)$), which was used, e.g., in [29] to show that the extension complexity of the correlation polytope is exponential. The clique number of $\mathrm{RG}(A)$ is also known as the *fooling set number* (see, e.g., [28]). Observe that the above combinatorial lower bounds only depend on the sparsity pattern of the matrix $A$, and that they are all equal to one for a strictly positive matrix.

Fawzi and Parrilo [27] have furthermore shown that the bound $\tau_+(A)$ is at least as good as norm-based lower bounds:

$$\tau_+(A) = \sup_{\substack{\mathscr{N} \text{ monotone and} \\ \text{positively homogeneous}}} \frac{\mathscr{N}^*(A)}{\mathscr{N}(A)}.$$

Here, a function $\mathscr{N} : \mathbb{R}_+^{m \times n} \to \mathbb{R}_+$ is *positively homogeneous* if $\mathscr{N}(\lambda A) = \lambda \mathscr{N}(A)$ for all $\lambda \geq 0$ and *monotone* if $\mathscr{N}(A) \leq \mathscr{N}(B)$ for $A \leq B$, and $\mathscr{N}^*(A)$ is defined as

$$\mathscr{N}^*(A) = \max\{L(A) : L : \mathbb{R}^{m \times n} \to \mathbb{R} \text{ linear and } L(X) \le 1 \text{ for all } X \in \mathbb{R}_+^{m \times n}$$
$$\text{with } \operatorname{rank}(X) \le 1 \text{ and } \mathscr{N}(X) \le 1\}.$$

These bounds are called *norm-based* since norms often provide valid functions $\mathscr{N}$. For example, when $\mathscr{N}$ is the $\ell_\infty$-norm, Rothvoß [66] used the corresponding lower bound to show that the matching polytope has exponential extension complexity.

When $\mathscr{N}$ is the Frobenius norm: $\mathscr{N}(A) = (\sum_{i,j} A_{ij}^2)^{1/2}$, the parameter $\mathscr{N}^*(A)$ is known as the *nonnegative nuclear norm*. In [26] it is denoted by $\nu_+(A)$, shown to satisfy $\operatorname{rank}_+(A) \ge (\nu_+(A)/||A||_F)^2$, and reformulated as

$$\nu_+(A) = \min\left\{\sum_i \lambda_i : A = \sum_i \lambda_i u_i v_i^\mathsf{T}, \; (\lambda_i, u_i, v_i) \in \mathbb{R}_+^{1+m+n}, \; ||u_i||_2 = ||v_i||_2 = 1\right\} \tag{26}$$

$$= \max\{\langle A, W \rangle : W \in \mathbb{R}^{m \times n}, \; \begin{pmatrix} I & -W \\ -W^\mathsf{T} & I \end{pmatrix} \text{ is copositive}\}. \tag{27}$$

where the cone of copositive matrices is the dual of the cone of completely positive matrices. Fawzi and Parrilo [26] use the copositive formulation (27) to provide bounds $\nu_+^{[k]}(A)$ ($k \ge 0$), based on inner approximations of the copositive cone from [60], which converge to $\nu_+(A)$ from below. We now observe that by Theorem 7 the atomic formulation of $\nu_+(A)$ from (26) can be seen as a moment optimization problem:

$$\nu_+(A) = \min \int_{V(S)} 1 \, \mathrm{d}\mu(x) \quad \text{s.t.} \quad A_{ij} = \int_{V(S)} x_i x_{m+j} \, \mathrm{d}\mu(x) \quad \text{for} \quad i \in [m], \, j \in [n].$$

Here, the optimization variable $\mu$ is required to be a Borel measure on the variety $V(S)$, where

$$S = \left\{\sum_{i=1}^m x_i^2 - 1, \; \sum_{j=1}^n x_{m+j}^2 - 1\right\}.$$

(The same observation is made in [74] for the real nuclear norm of a symmetric 3-tensor and in [59] for symmetric odd-dimensional tensors.) For $t \in \mathbb{N} \cup \{\infty\}$, let $\mu_t(A)$ denote the parameter defined analogously to $\xi_t^+(A)$, where we replace the condition $L \ge 0$ on $\mathscr{M}_{2t}(S_A^+)$ by $L \ge 0$ on $\mathscr{M}_{2t}(\{x_1, \ldots, x_{m+n}\})$ and $L = 0$ on $\mathscr{I}_{2t}(S)$, and let $\mu_*(A)$ be obtained by adding the constraint $\operatorname{rank}(M(L)) < \infty$ to $\mu_\infty(A)$. We have $\mu_t(A) \to \mu_\infty(A) = \mu_*(A) = \nu_+(A)$ by Theorem 7 and (a non-normalized analog of) Theorem 8. One can show that $\mu_1(A)$ with the additional constraints $L(u) \ge 0$ for all $u \in [\mathbf{x}]_2$, is at least as good as $\nu_+^{[0]}(A)$. It is not clear how the hierarchies $\mu_t(A)$ and $\nu_+^{[k]}(A)$ compare in general.

## 4.2 Computational Examples

We illustrate the performance of our approach by comparing our lower bounds $\xi_{2,\dagger}^+$ and $\xi_{3,\dagger}^+$ to the lower bounds $\tau_+$ and $\tau_+^{\text{sos}}$ on the two examples considered in [27].

### 4.2.1 All Nonnegative 2 × 2 Matrices

For $A(\alpha) = \left(\begin{smallmatrix} 1 & 1 \\ 1 & \alpha \end{smallmatrix}\right)$, Fawzi and Parrilo [27] show that

$$\tau_+(A(\alpha)) = 2 - \alpha \quad \text{and} \quad \tau_+^{\text{sos}}(A(\alpha)) = \frac{2}{1+\alpha} \quad \text{for all} \quad 0 \le \alpha \le 1.$$

Since the parameters $\tau_+(A)$ and $\tau_+^{\text{sos}}(A)$ are invariant under scaling and permuting rows and columns of $A$, one can use the identity

$$\begin{pmatrix} 1 & 1 \\ 1 & \alpha \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1/\alpha \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

to see this describes the parameters for all nonnegative $2 \times 2$ matrices. By using a semidefinite programming solver for $\alpha = k/100$, $k \in [100]$, we see that $\xi_2^+(A(\alpha))$ coincides with $\tau_+(A(\alpha))$.

### 4.2.2 The Nested Rectangles Problem

In this section, we consider the nested rectangles problem as described in [27, Section 2.7.2] (see also [55]), which asks for which $a, b$ there exists a triangle $T$ such that $R(a, b) \subseteq T \subseteq P$, where $R(a, b) = [-a, a] \times [-b, b]$ and $P = [-1, 1]^2$.

The nonnegative rank relates not only to the extension complexity of a polytope [78], but also to extended formulations of nested pairs [12,31]. An *extended formulation* of a pair of polytopes $P_1 \subseteq P_2 \subseteq \mathbb{R}^d$ is a (possibly) higher dimensional polytope $K$ whose projection $\pi(K)$ is nested between $P_1$ and $P_2$. Let us suppose $\pi(K) = \{x \in \mathbb{R}^d : y \in \mathbb{R}_+^k, (x, y) \in K\}$ and $K = \{(x, y) : Ex + Fy = g, y \in \mathbb{R}_+^k\}$, then $k$ is the *size of the extended formulation*, and the smallest such $k$ is called the *extension complexity* of the pair $(P_1, P_2)$. It is known (cf. [12, Theorem 1]) that the extension complexity of the pair $(P_1, P_2)$, where

$$P_1 = \text{conv}(\{v_1, \ldots, v_n\}) \quad \text{and} \quad P_2 = \left\{ x : a_i^\mathsf{T} x \le b_i \text{ for } i \in [m] \right\},$$
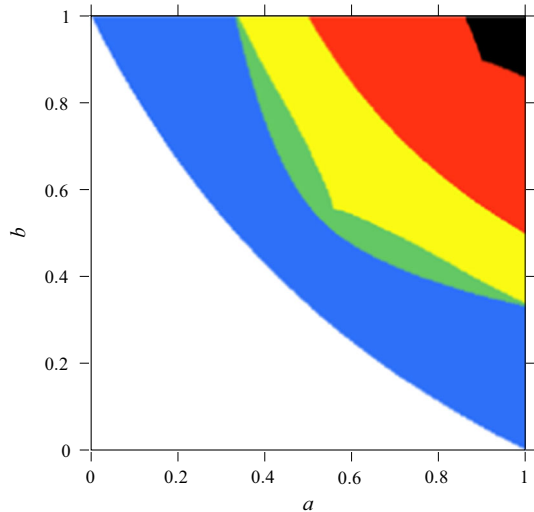
is equal to the nonnegative rank of the generalized slack matrix $S_{P_1, P_2} \in \mathbb{R}^{m \times n}$, defined by

$$(S_{P_1, P_2})_{ij} = b_j - a_j^\mathsf{T} v_i \quad \text{for} \quad i \in [m], j \in [n].$$

Any nonnegative matrix is the slack matrix of some nested pair of polytopes [34, Lemma 4.1] (see also [31]).

Applying this to the pair $(R(a, b), P)$, one immediately sees that there exists a polytope $K$ with at most three facets whose projection $T = \pi(K) \subseteq \mathbb{R}^2$ satisfies $R(a, b) \subseteq T \subseteq P$ if and only if the pair $(R(a, b), P)$ admits an extended formulation of size 3. For $a, b > 0$, the polytope $T$ has to be 2 dimensional, therefore $K$ has to be at least 2 dimensional as well; it follows that $K$ and $T$ have to be triangles. Hence,

**Fig. 1** The colored region corresponds to $\mathrm{rank}_+(S(a,b)) = 4$. The top right region (black) corresponds to $\xi^+_{1,\dagger}(S(a,b)) > 3$, the two top right regions (black and red) together correspond to $\tau^{\mathrm{sos}}_+(S(a,b)) > 3$, the three top right regions (black, red and yellow) to $\xi^+_{2,\dagger}(S(a,b)) > 3$, and the four top right regions (black, red, yellow, and green) to $\xi^+_{3,\dagger}(S(a,b)) > 3$ (Color figure online)

there exists a triangle $T$ such that $R(a,b) \subseteq T \subseteq P$ if and only if the nonnegative rank of the slack matrix $S(a,b) := S_{R(a,b),P}$ is equal to 3. One can verify that

$$S(a,b) = \begin{pmatrix} 1-a & 1+a & 1-b & 1+b \\ 1+a & 1-a & 1-b & 1+b \\ 1+a & 1-a & 1+b & 1-b \\ 1-a & 1+a & 1+b & 1-b \end{pmatrix}.$$

Such a triangle exists if and only if $(1+a)(1+b) \leq 2$ (see [27, Proposition 4] for a proof sketch). To test the quality of their bound, Fawzi and Parrilo [27] compute $\tau^{\mathrm{sos}}_+(S(a,b))$ for different values of $a$ and $b$. In doing so, they determine the region where $\tau^{\mathrm{sos}}_+(S(a,b)) > 3$. We do the same for the bounds $\xi^+_{1,\dagger}(S(a,b))$, $\xi^+_{2,\dagger}(S(a,b))$ and $\xi^+_{3,\dagger}(S(a,b))$, see Fig. 1. The results show that $\xi^+_{2,\dagger}(S(a,b))$ strictly improves upon the bound $\tau^{\mathrm{sos}}_+(S(a,b))$, and that $\xi^+_{3,\dagger}(S(a,b))$ is again a strict improvement over $\xi^+_{2,\dagger}(S(a,b))$.

## 5 Lower Bounds on the Positive Semidefinite Rank

The positive semidefinite rank can be seen as an *asymmetric* version of the completely positive semidefinite rank. Hence, as was the case in the previous section for the nonnegative rank, we need to select suitable factors in a minimal factorization in order to be able to bound their maximum eigenvalues and obtain a localizing set of polynomials leading to an Archimedean quadratic module.

For this we can follow, e.g., the approach in [52, Lemma 5] to rescale a factorization and claim that, for any $A \in \mathbb{R}^{m \times n}_+$ with psd-$\mathrm{rank}_\mathbb{C}(A) = d$, there exists a factorization $A = (\langle X_i, X_{m+j} \rangle)$ by matrices $X_1, \ldots, X_{m+n} \in \mathrm{H}^d_+$ such that $\sum_{i=1}^m X_i = I$ and

$\mathrm{Tr}(X_{m+j}) = \sum_i A_{ij}$ for all $j \in [n]$. Indeed, starting from any factorization $X_i$, $X_{m+j}$ in $\mathrm{H}_+^d$ of $A$, we may replace $X_i$ by $X^{-1/2} X_i X^{-1/2}$ and $X_{m+j}$ by $X^{1/2} X_{m+j} X^{1/2}$, where $X := \sum_{i=1}^m X_i$ is positive definite (by minimality of $d$). This argument shows that the set of polynomials

$$S_A^{\mathrm{psd}} = \left\{ x_i - x_i^2 : i \in [m] \right\} \cup \left\{ \left( \sum_{i=1}^m A_{ij} \right) x_{m+j} - x_{m+j}^2 : j \in [n] \right\}$$

is localizing for $A$; that is, there is *at least one* minimal factorization **X** of $A$ such that $g(\mathbf{X}) \succeq 0$ for all polynomials $g \in S_A^{\mathrm{psd}}$. Moreover, for the same minimal factorization **X** of $A$, we have $p(\mathbf{X})(1 - \sum_{i=1}^m X_i) = 0$ for all $p \in \mathbb{R}\langle \mathbf{x} \rangle$.

Given $A \in \mathbb{R}_+^{m \times n}$, for each $t \in \mathbb{N} \cup \{\infty\}$ we consider the semidefinite program

$$\begin{aligned}
\xi_t^{\mathrm{psd}}(A) = \min \big\{ L(1) : \; & L \in \mathbb{R}\langle x_1, \ldots, x_{m+n} \rangle_{2t}^*, \\
& L(x_i x_{m+j}) = A_{ij} \quad \text{for} \quad i \in [m], j \in [n], \\
& L \geq 0 \quad \text{on} \quad \mathcal{M}_{2t}(S_A^{\mathrm{psd}}), \\
& L = 0 \quad \text{on} \quad \mathcal{I}_{2t}(1 - \textstyle\sum_{i=1}^m x_i) \big\}.
\end{aligned}$$

We additionally define $\xi_*^{\mathrm{psd}}(A)$ by adding the constraint $\mathrm{rank}(M(L)) < \infty$ to the program defining $\xi_\infty^{\mathrm{psd}}(A)$ (and considering the infimum instead of the minimum, since we do not know if the infimum is attained in $\xi_*^{\mathrm{psd}}(A)$). By the above discussion, it follows that the parameter $\xi_*^{\mathrm{psd}}(A)$ is a lower bound on psd-rank$_\mathbb{C}(A)$ and we have

$$\xi_1^{\mathrm{psd}}(A) \leq \ldots \leq \xi_t^{\mathrm{psd}}(A) \leq \ldots \leq \xi_\infty^{\mathrm{psd}}(A) \leq \xi_*^{\mathrm{psd}}(A) \leq \text{psd-rank}_\mathbb{C}(A).$$

Note that, in contrast to the previous bounds, the parameter $\xi_t^{\mathrm{psd}}(A)$ is not invariant under rescaling the rows of $A$ or under taking the transpose of $A$ (see Sect. 5.2.2).

It follows from the construction of $S_A^{\mathrm{psd}}$ and Eq. (10) that the quadratic module $\mathcal{M}(S_A^{\mathrm{psd}})$ is Archimedean, and hence the following analog of Proposition 1 can be shown.

**Proposition 16** *Let $A \in \mathbb{R}_+^{m \times n}$. For each $t \in \mathbb{N} \cup \{\infty\}$, the optimum in $\xi_t^{\mathrm{psd}}(A)$ is attained, and we have*

$$\lim_{t \to \infty} \xi_t^{\mathrm{psd}}(A) = \xi_\infty^{\mathrm{psd}}(A).$$

*Moreover, $\xi_\infty^{\mathrm{psd}}(A)$ is equal to the infimum over all $\alpha \geq 0$ for which there exists a unital $C^*$-algebra $\mathcal{A}$ with tracial state $\tau$ and $\mathbf{X} \in \mathcal{D}_\mathcal{A}(S_A^{\mathrm{psd}}) \cap \mathcal{V}_\mathcal{A}(1 - \sum_{i=1}^m x_i)$ such that $A = \alpha \cdot (\tau(X_i X_{m+j}))_{i \in [m], j \in [n]}$.*

### 5.1 Comparison to Other Bounds

In [52], the following bound on the complex positive semidefinite rank was derived:

$$\text{psd-rank}_{\mathbb{C}}(A) \geq \sum_{i=1}^{m} \max_{j \in [n]} \frac{A_{ij}}{\sum_i A_{ij}}. \tag{28}$$

If a feasible linear form $L$ to $\xi_t^{\text{psd}}(A)$ satisfies the inequalities $L(x_i(\sum_i A_{ij} - x_{m+j})) \geq 0$ for all $i \in [m]$, $j \in [n]$, then $L(1)$ is at least the above lower bound. Indeed, the inequalities give

$$L(x_i) \geq \max_{j \in [n]} \frac{L(x_i x_{m+j})}{\sum_i A_{ij}} = \max_{j \in [n]} \frac{A_{ij}}{\sum_i A_{ij}}.$$

and hence

$$L(1) = \sum_{i=1}^{m} L(x_i) \geq \sum_{i=1}^{m} \max_{j \in [n]} \frac{A_{ij}}{\sum_i A_{ij}}.$$

The inequalities $L(x_i(\sum_i A_{ij} - x_{m+j})) \geq 0$ are easily seen to be valid for trace evaluations at points of $\mathscr{D}(S_A^{\text{psd}})$. More importantly, as in Lemma 2, these inequalities are satisfied by feasible linear forms to the programs $\xi_\infty^{\text{psd}}(A)$ and $\xi_*^{\text{psd}}(A)$. Hence, $\xi_\infty^{\text{psd}}(A)$ and $\xi_*^{\text{psd}}(A)$ are at least as good as the lower bound (28).

In [52], two other fidelity based lower bounds on the psd-rank were defined; we do not know how they compare to $\xi_t^{\text{psd}}(A)$.

### 5.2 Computational Examples

In this section, we apply our bounds to some (small) examples taken from the literature, namely $3 \times 3$ circulant matrices and slack matrices of small polygons.

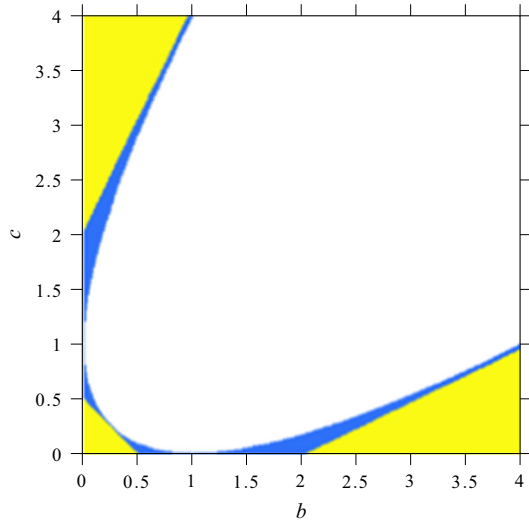#### 5.2.1 Nonnegative Circulant Matrices of Size 3

We consider the nonnegative circulant matrices of size 3 which are, up to scaling, of the form

$$M(b, c) = \begin{pmatrix} 1 & b & c \\ c & 1 & b \\ b & c & 1 \end{pmatrix} \quad \text{with} \quad b, c \geq 0.$$

If $b = 1 = c$, then $\text{rank}(M(b, c)) = \text{psd-rank}_{\mathbb{R}}(M(b, c)) = \text{psd-rank}_{\mathbb{C}}(M(b, c)) = 1$. Otherwise, we have $\text{rank}(M(b, c)) \geq 2$, which implies $\text{psd-rank}_{\mathbb{K}}(M(b, c)) \geq 2$ for $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. In [25, Example 2.7] it is shown that

$$\text{psd-rank}_{\mathbb{R}}(M(b, c)) \leq 2 \quad \Longleftrightarrow \quad 1 + b^2 + c^2 \leq 2(b + c + bc).$$

Hence, if $b$ and $c$ do not satisfy the above relation then psd-rank$_\mathbb{R}(M(b, c)) = 3$.

To see how good our lower bounds are for this example, we use a semidefinite programming solver to compute $\xi_2^{\text{psd}}(M(b, c))$ for $(b, c) \in [0, 4]^2$ (with stepsize $0.01$). In Fig. 2, we see that the bound $\xi_2^{\text{psd}}(M(b, c))$ certifies that psd-rank$_\mathbb{R}(M(b, c)) =$ psd-rank$_\mathbb{C}(M(b, c)) = 3$ for most values $(b, c)$ where psd-rank$_\mathbb{R}(M(b, c)) = 3$.

### 5.2.2 Polygons

Here, we consider the slack matrices of two polygons in the plane, where the bounds are sharp (after rounding) and illustrate the dependence on scaling the rows or taking the transpose. We consider the quadrilateral $Q$ with vertices $(0, 0)$, $(0, 1)$, $(1, 0)$, $(2, 2)$, and the regular hexagon $H$, whose slack matrices are given by

$$S_Q = \begin{pmatrix} 0 & 0 & 2 & 2 \\ 1 & 0 & 0 & 3 \\ 0 & 1 & 3 & 0 \\ 2 & 2 & 0 & 0 \end{pmatrix}, \qquad S_H = \begin{pmatrix} 0 & 1 & 2 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 \end{pmatrix}.$$

Our lower bounds on the psd-rank$_\mathbb{C}$ are not invariant under taking the transpose, indeed numerically we have $\xi_2^{\text{psd}}(S_Q) \approx 2.266$ and $\xi_2^{\text{psd}}(S_Q^\mathsf{T}) \approx 2.5$. The slack matrix $S_Q$ has psd-rank$_\mathbb{R}(S_Q) = 3$ (a corollary of [35, Theorem 4.3]) and therefore both bounds certify psd-rank$_\mathbb{C}(S_Q) = 3 = $ psd-rank$_\mathbb{R}(S_Q)$.

Secondly, our bounds are not invariant under rescaling the rows of a nonnegative matrix. Numerically we have $\xi_2^{\text{psd}}(S_H) \approx 1.99$ while $\xi_2^{\text{psd}}(DS_H) \approx 2.12$, where $D = \text{Diag}(2, 2, 1, 1, 1, 1)$. The bound $\xi_2^{\text{psd}}(DS_H)$ is in fact tight (after rounding) for

the complex positive semidefinite rank of $DS_H$ and hence of $S_H$: in [33] it is shown that psd-rank$_{\mathbb{C}}(S_H) = 3$.

## 6 Discussion and Future Work

In this work, we provide a unified approach for the four matrix factorizations obtained by considering (a)symmetric factorizations by nonnegative vectors and positive semidefinite matrices. Our methods can be extended to the nonnegative tensor rank, which is defined as the smallest integer $d$ for which a $k$-tensor $A \in \mathbb{R}_+^{n_1 \times \cdots \times n_k}$ can be written as $A = \sum_{l=1}^d u_{1,l} \otimes \cdots \otimes u_{k,l}$ for nonnegative vectors $u_{j,l} \in \mathbb{R}_+^{n_j}$. The approach from Sect. 4 for rank$_+$ can be extended to obtain a hierarchy of lower bounds on the nonnegative tensor rank. For instance, if $A$ is a 3-tensor, the analogous bound $\xi_+^+(A)$ is obtained by minimizing $L(1)$ over $L \in \mathbb{R}[x_1, \ldots, x_{n_1+n_2+n_3}]^*$ such that $L(x_{i_1} x_{n_1+i_2} x_{n_1+n_2+i_3}) = A_{i_1 i_2 i_3}$ (for $i_1 \in [n_1]$, $i_2 \in [n_2]$, $i_3 \in [n_3]$), using as localizing polynomials in $S_A^+$ the polynomials $\sqrt[3]{A_{\max}} x_i - x_i^2$ and $A_{i_1 i_2 i_3} - x_{i_1} x_{n_1+i_2} x_{n_1+n_2+i_3}$. As in the matrix case one can compare to the bounds $\tau_+(A)$ and $\tau_+^{\mathrm{sos}}(A)$ from [27]. One can show $\xi_*^+(A) = \tau_+(A)$, and one can show $\xi_{3,\dagger}^+(A) \geq \tau_+^{\mathrm{sos}}(A)$ after adding the conditions $L(x_{i_1} x_{n_1+i_2} x_{n_1+n_2+i_3}(A_{i_1 i_2 i_3} - x_{i_1} x_{n_1+i_2} x_{n_1+n_2+i_3})) \geq 0$ to $\xi_3^+(A)$.

Testing membership in the completely positive cone and the completely positive semidefinite cone is another important problem, to which our hierarchies can also be applied. It follows from the proof of Proposition 8 that if $A$ is not completely positive then, for some order $t$, the program $\xi_t^{\mathrm{cp}}(A)$ is infeasible or its optimum value is larger than the Caratheodory bound on the cp-rank (which is similar to an earlier result in [58]). In the noncommutative setting, the situation is more complicated: If $\xi^{\mathrm{cpsd}}(A)$ is feasible, then $A \in \mathrm{CS}_+$, and if $A \notin \mathrm{CS}_{+,\mathrm{vN}}^n$, then $\xi_\infty^{\mathrm{cpsd}}(A)$ is infeasible (Propositions 1 and 2). Here, $\mathrm{CS}_{+,\mathrm{vN}}^n$ is the cone defined in [17] consisting of the matrices admitting a factorization in a von Neumann algebra with a trace. By Lemma 12, $\mathrm{CS}_{+,\mathrm{vN}}^n$ can equivalently be characterized as the set of matrices of the form $\alpha \, (\tau(a_i a_j))$ for some $C^*$-algebra $\mathscr{A}$ with tracial state $\tau$, positive elements $a_1, \ldots, a_n \in \mathscr{A}$ and $\alpha \in \mathbb{R}_+$.

Our lower bounds are on the *complex* version of the (completely) positive semidefinite rank. As far as we are aware, the existing lower bounds (except for the dimension counting rank lower bound) are also on the complex (completely) positive semidefinite rank. It would be interesting to find a lower bound on the *real* (completely) positive semidefinite rank that can go beyond the complex (completely) positive semidefinite rank.

We conclude with some open questions regarding applications of lower bounds on matrix factorization ranks. First, as was shown in [38,62,63], completely positive semidefinite matrices whose cpsd-rank$_{\mathbb{C}}$ is larger than their size do exist, but currently we do not know how to construct small examples for which this holds. Hence, a concrete question: Does there exist a $5 \times 5$ completely positive semidefinite matrix whose cpsd-rank$_{\mathbb{C}}$ is at least 6? Second, as we mentioned before, the asymmetric setting corresponds to (semidefinite) extension complexity of polytopes. Rothvoß' result [66] (indirectly) shows that the parameter $\xi_\infty^+$ is exponential (in the number of

nodes of the graph) for the slack matrix of the matching polytope. Can this result also be shown directly using the dual formulation of $\xi_\infty^+$, that is, by a sum-of-squares certificate? If so, could one extend the argument to the noncommutative setting (which would show a lower bound on the semidefinite extension complexity)?

## A Commutative and Tracial Polynomial Optimization

In this appendix, we discuss known convergence and flatness results for commutative and tracial polynomial optimization. We present these results in such a way that they can be directly used for our hierarchies of lower bounds on matrix factorization ranks. Although the commutative case was developed first, here we treat the commutative and tracial cases together. For the reader's convenience, we provide all proofs by working on the "moment side"; that is, relying on properties of linear functionals rather than using real algebraic results on sums of squares. Tracial optimization is an adaptation of eigenvalue optimization as developed in [61], but here we only discuss the commutative and tracial cases, as these are most relevant to our work.

### A.1 Flat Extensions and Representations of Linear Forms

The optimization variables in the optimization problems considered in this paper are linear forms on spaces of (noncommutative) polynomials. To study the properties of the bounds obtained through these optimization problems, we need to study properties and representations of (flat) linear forms on polynomial spaces.

In Sect. 1.3, the key examples of symmetric tracial linear functionals on $\mathbb{R}\langle\mathbf{x}\rangle_{2t}$ are trace evaluations on a (finite dimensional) $C^*$-algebra. In this section, we present some results that provide conditions under which, conversely, a symmetric tracial linear map on $\mathbb{R}\langle\mathbf{x}\rangle_{2t}$ ($t \in \mathbb{N} \cup \{\infty\}$) that is nonnegative on $\mathcal{M}(S)$ and zero on $\mathcal{I}(T)$ arises from trace evaluations at elements in the intersection of the $C^*$-algebraic analogs of the matrix positivity domain of $S$ and the matrix ideal of $T$. In Theorems 1 and 2, we consider the case $t = \infty$ and in Theorem 3 we consider the case $t \in \mathbb{N}$. Results like these can for instance be used to link the linear forms arising in the limiting optimization problems of our hierarchies to matrix factorization ranks.

The proofs of Theorems 1 and 2 use a classical Gelfand–Naimark–Segal (GNS) construction. In these proofs, it will also be convenient to work with the concept of the null space of a linear functional $L \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}^*$, which is defined as the vector space

$$N_t(L) = \big\{ p \in \mathbb{R}\langle\mathbf{x}\rangle_t : L(qp) = 0 \text{ for } q \in \mathbb{R}\langle\mathbf{x}\rangle_t \big\}.$$

We use the notation $N(L) = N_\infty(L)$ for the nontruncated null space. Recall that $M_t(L)$ is the moment matrix associated with $L$, its rows and columns are indexed by words in $\langle \mathbf{x} \rangle_t$, and its entries are given by $M_t(L)_{w,w'} = L(w^*w')$ for $w, w' \in \langle \mathbf{x} \rangle_t$. The null space of $L$ can therefore be identified with the kernel of $M_t(L)$: A polynomial $p = \sum_w c_w w$ belongs to $N_t(L)$ if and only if its coefficient vector $(c_w)$ belongs to the kernel of $M_t(L)$.

In Sect. 1.3, we defined a linear functional $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ to be $\delta$-flat based on the rank stabilization property (4) of its moment matrix: $\operatorname{rank}(M_t(L)) = \operatorname{rank}(M_{t-\delta}(L))$. This definition can be reformulated in terms of a decomposition of the corresponding polynomial space using the null space: the form $L$ is $\delta$-flat if and only if

$$\mathbb{R}\langle \mathbf{x} \rangle_t = \mathbb{R}\langle \mathbf{x} \rangle_{t-\delta} + N_t(L).$$

Recall that $L$ is said to be flat if it is $\delta$-flat for some $\delta \geq 1$. Finally, in the nontruncated case ($t = \infty$), $L$ was called flat if $\operatorname{rank}(M(L)) < \infty$. We can now see that $\operatorname{rank}(M(L)) < \infty$ if and only if there exists an integer $s \in \mathbb{N}$ such that $\mathbb{R}\langle \mathbf{x} \rangle = \mathbb{R}\langle \mathbf{x} \rangle_s + N(L)$.

Theorem 1 below is implicit in several works (see, e.g., [16,57]). Here, we assume that $\mathcal{M}(S) + \mathcal{I}(T)$ is Archimedean, which we recall means that there exists a scalar $R > 0$ such that

$$R - \sum_{i=1}^n x_i^2 \in \mathcal{M}(S) + \mathcal{I}(T).$$

**Theorem 1** *Let* $S \subseteq \operatorname{Sym} \mathbb{R}\langle \mathbf{x} \rangle$ *and* $T \subseteq \mathbb{R}\langle \mathbf{x} \rangle$ *with* $\mathcal{M}(S) + \mathcal{I}(T)$ *Archimedean. Given a linear form* $L \in \mathbb{R}\langle \mathbf{x} \rangle^*$, *the following are equivalent:*

(1) *$L$ is symmetric, tracial, nonnegative on $\mathcal{M}(S)$, zero on $\mathcal{I}(T)$, and $L(1) = 1$;*
(2) *there is a unital $C^*$-algebra $\mathscr{A}$ with tracial state $\tau$ and $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S) \cap \mathscr{V}_{\mathscr{A}}(T)$ with*

$$L(p) = \tau(p(\mathbf{X})) \quad \text{for all} \quad p \in \mathbb{R}\langle \mathbf{x} \rangle. \tag{29}$$

**Proof** We first prove the easy direction (2) $\Rightarrow$ (1): We have

$$L(p^*) = \tau(p^*(\mathbf{X})) = \tau(p(\mathbf{X})^*) = \overline{\tau(p(\mathbf{X}))} = \overline{L(p)} = L(p),$$

where we use that $\tau$ is Hermitian and $X_i^* = X_i$ for $i \in [n]$. Moreover, $L$ is tracial since $\tau$ is tracial. In addition, for $g \in S \cup \{1\}$ and $p \in \mathbb{R}\langle \mathbf{x} \rangle$ we have

$$L(p^*gp) = \tau(p^*(\mathbf{X})g(\mathbf{X})p(\mathbf{X})) = \tau(p(\mathbf{X})^*g(\mathbf{X})p(\mathbf{X})) \geq 0,$$

since $g(\mathbf{X})$ is positive in $\mathscr{A}$ as $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S)$ and $\tau$ is positive. Similarly $L(hq) = \tau(h(\mathbf{X})q(\mathbf{X})) = 0$ for all $h \in T$, since $\mathbf{X} \in \mathscr{V}_{\mathscr{A}}(\mathbf{T})$.

We show (1) $\Rightarrow$ (2) by applying a GNS construction. Consider the quotient vector space $\mathbb{R}\langle \mathbf{x} \rangle / N(L)$, and denote the class of $p$ in $\mathbb{R}\langle \mathbf{x} \rangle / N(L)$ by $\overline{p}$. We can equip this quotient with the inner product $\langle \overline{p}, \overline{q} \rangle = L(p^*q)$ for $p, q \in \mathbb{R}\langle \mathbf{x} \rangle$, so that the

completion $\mathscr{H}$ of $\mathbb{R}\langle\mathbf{x}\rangle/N(L)$ is a separable Hilbert space. As $N(L)$ is a left ideal in $\mathbb{R}\langle\mathbf{x}\rangle$, the operator

$$X_i : \mathbb{R}\langle\mathbf{x}\rangle/N(L) \to \mathbb{R}\langle\mathbf{x}\rangle/N(L), \ \overline{p} \mapsto \overline{x_i p} \tag{30}$$

is well defined. We have

$$\langle X_i \, \overline{p}, \overline{q}\rangle = L((x_i p)^* q) = L(p^* x_i q) = \langle \overline{p}, X_i \overline{q}\rangle \quad \text{for all} \quad p, q \in \mathbb{R}\langle\mathbf{x}\rangle,$$

so the $X_i$ are self-adjoint. Since $g \in S \cup \{1\}$ is symmetric and $\langle \overline{p}, g(\mathbf{X})\overline{p}\rangle = \langle \overline{p}, \overline{gp}\rangle = L(p^* g p) \geq 0$ for all $p$ we have $g(\mathbf{X}) \succeq 0$. By the Archimedean condition, there exists an $R > 0$ such that $R - \sum_{i=1}^{n} x_i^2 \in \mathscr{M}(S) + \mathscr{I}(T)$. Using $R - x_i^2 = (R - \sum_{j=1}^{n} x_j^2) + \sum_{j \neq i} x_j^2 \in \mathscr{M}(S) + \mathscr{I}(T)$, we get

$$\langle X_i \overline{p}, X_i \overline{p}\rangle = L(p^* x_i^2 p) \leq R \cdot L(p^* p) = R \langle \overline{p}, \overline{p}\rangle \quad \text{for all} \quad p \in \mathbb{R}\langle\mathbf{x}\rangle.$$

So each $X_i$ extends to a bounded self-adjoint operator, also denoted $X_i$, on the Hilbert space $\mathscr{H}$ such that $g(\mathbf{X})$ is positive for all $g \in S \cup \{1\}$. Moreover, we have $\langle \overline{f}, h(\mathbf{X})\overline{1}\rangle = L(f^* h) = 0$ for all $f \in \mathbb{R}\langle\mathbf{x}\rangle, h \in T$.

The operators $X_i \in \mathscr{B}(\mathscr{H})$ extend to self-adjoint operators in $\mathscr{B}(\mathbb{C} \otimes_{\mathbb{R}} \mathscr{H})$, where $\mathbb{C} \otimes_{\mathbb{R}} \mathscr{H}$ is the complexification of $\mathscr{H}$. Let $\mathscr{A}$ be the unital $C^*$-algebra obtained by taking the operator norm closure of $\mathbb{R}\langle\mathbf{X}\rangle \subseteq \mathscr{B}(\mathbb{C} \otimes_{\mathbb{R}} \mathscr{H})$. It follows that $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S) \cap \mathscr{V}_{\mathscr{A}}(T)$.

Define the state $\tau$ on $\mathscr{A}$ by $\tau(a) = \langle \overline{1}, a\overline{1}\rangle$ for $a \in \mathscr{A}$. For all $p, q \in \mathbb{R}\langle\mathbf{x}\rangle$, we have

$$\tau(p(\mathbf{X})q(\mathbf{X})) = \langle \overline{1}, p(\mathbf{X})q(\mathbf{X})\overline{1}\rangle = \langle \overline{1}, \overline{pq}\rangle = L(pq), \tag{31}$$

so that the restriction of $\tau$ to $\mathbb{R}\langle\mathbf{X}\rangle$ is tracial. Since $\mathbb{R}\langle\mathbf{X}\rangle$ is dense in $\mathscr{A}$ in the operator norm, this implies $\tau$ is tracial.

To conclude the proof, observe that (29) follows from (31) by taking $q = 1$. ☐

The next result can be seen as a finite dimensional analog of the above result, where we do not need $\mathscr{M}(S) + \mathscr{I}(T)$ to be Archimedean, but instead we assume the rank of $M(L)$ to be finite (i.e., $L$ to be flat). In addition to the Gelfand–Naimark–Segal construction, the proof uses Artin–Wedderburn theory. For the unconstrained case, the proof of this result can be found in [15], and in [16,43] this result is extended to the constrained case.

**Theorem 2** *For $S \subseteq \operatorname{Sym} \mathbb{R}\langle\mathbf{x}\rangle$, $T \subseteq \mathbb{R}\langle\mathbf{x}\rangle$, and $L \in \mathbb{R}\langle\mathbf{x}\rangle^*$, the following are equivalent:*

(1) *$L$ is a symmetric, tracial, linear form with $L(1) = 1$ that is nonnegative on $\mathscr{M}(S)$, zero on $\mathscr{I}(T)$, and has $\operatorname{rank}(M(L)) < \infty$;*
(2) *there is a finite dimensional $C^*$-algebra $\mathscr{A}$ with a tracial state $\tau$, and $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S) \cap \mathscr{V}_{\mathscr{A}}(T)$ satisfying equation (29);*
(3) *$L$ is a convex combination of normalized trace evaluations at points in $\mathscr{D}(S) \cap \mathscr{V}(T)$.*

**Proof** $((1) \Rightarrow (2))$ Here, we can follow the proof of Theorem 1, with the extra observation that the condition $\mathrm{rank}(M(L)) < \infty$ implies that the quotient space $\mathbb{R}\langle \mathbf{x} \rangle / N(L)$ is finite dimensional. Since $\mathbb{R}\langle \mathbf{x} \rangle / N(L)$ is finite dimensional the multiplication operators are bounded, and the constructed $C^*$-algebra $\mathscr{A}$ is finite dimensional.

$((2) \Rightarrow (3))$ By Artin-Wedderburn theory, there exists a $*$-isomorphism

$$\varphi \colon \mathscr{A} \to \bigoplus_{m=1}^{M} \mathbb{C}^{d_m \times d_m} \quad \text{for some } M \in \mathbb{N}, \ d_1, \ldots, d_M \in \mathbb{N}.$$

Define the $*$-homomorphisms $\varphi_m \colon \mathscr{A} \to \mathbb{C}^{d_m \times d_m}$ for $m \in [M]$ by $\varphi = \oplus_{m=1}^{M} \varphi_m$. Then, for each $m \in [M]$, the map $\mathbb{C}^{d_m \times d_m} \to \mathbb{C}$ defined by $X \mapsto \tau(\varphi_m^{-1}(X))$ is a positive tracial linear form, and hence, it is a nonnegative multiple $\lambda_m \mathrm{tr}(\cdot)$ of the normalized matrix trace (since, for a full matrix algebra, the normalized trace is the unique tracial state). Then, we have $\tau(a) = \sum_m \lambda_m \, \mathrm{tr}(\varphi_m(a))$ for all $a \in \mathscr{A}$. So $\tau(\cdot) = \sum_m \lambda_m \mathrm{tr}(\cdot)$ for nonnegative scalars $\lambda_m$ with $\sum_m \lambda_m = L(1) = 1$. By defining the matrices $X_i^m = \varphi_m(X_i)$ for $m \in [M]$, we get

$$L(p) = \tau(p(X_1, \ldots, X_n)) = \sum_{m=1}^{M} \lambda_m \, \mathrm{tr}(p(X_1^m, \ldots, X_n^m)) \quad \text{for all} \quad p \in \mathbb{R}\langle \mathbf{x} \rangle.$$

Since $\varphi_m$ is a $*$-homomorphism, we have $g(X_1^m, \ldots, X_n^m) \succeq 0$ for all $g \in S \cup \{1\}$ and also $h(X_1^m, \ldots, X_n^m) = 0$ for all $h \in T$, which shows $(X_1^m, \ldots, X_n^m) \in \mathscr{D}(S) \cap \mathscr{V}(T)$.

$((3) \Rightarrow (1))$ If $L$ is a conic combination of trace evaluations at elements from $\mathscr{D}(S) \cap \mathscr{V}(T)$, then $L$ is symmetric, tracial, nonnegative on $\mathscr{M}(S)$, zero on $\mathscr{I}(T)$, and satisfies $\mathrm{rank}(M(L)) < \infty$ because the moment matrix of any trace evaluation has finite rank. □

The previous two theorems were about linear functionals defined on the full space of noncommutative polynomials. The following result claims that a *flat* linear functional on a truncated polynomial space can be extended to a flat linear functional on the full space of polynomials while preserving the same positivity properties. It is due to Curto and Fialkow [19] in the commutative case and extensions to the noncommutative case can be found in [61] (for eigenvalue optimization) and [15] (for trace optimization).

**Theorem 3** *Let* $1 \leq \delta \leq t < \infty$, $S \subseteq \mathrm{Sym}\, \mathbb{R}\langle \mathbf{x} \rangle_{2\delta}$, *and* $T \subseteq \mathbb{R}\langle \mathbf{x} \rangle_{2\delta}$. *Suppose* $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ *is symmetric, tracial, $\delta$-flat, nonnegative on* $\mathscr{M}_{2t}(S)$, *and zero on* $\mathscr{I}_{2t}(T)$. *Then* $L$ *extends to a symmetric, tracial, linear form on* $\mathbb{R}\langle \mathbf{x} \rangle$ *that is nonnegative on* $\mathscr{M}(S)$, *zero on* $\mathscr{I}(T)$, *and whose moment matrix has finite rank.*

**Proof** Let $W \subseteq \langle \mathbf{x} \rangle_{t-\delta}$ index a maximum nonsingular submatrix of $M_{t-\delta}(L)$, and let $\mathrm{span}(W)$ be the linear space spanned by $W$. We have the vector space direct sum

$$\mathbb{R}\langle \mathbf{x} \rangle_t = \mathrm{span}(W) \oplus N_t(L). \tag{32}$$

That is, for each $u \in \langle \mathbf{x} \rangle_t$ there exists a unique $r_u \in \mathrm{span}(W)$ such that $u - r_u \in N_t(L)$.

We first construct the (unique) symmetric flat extension $\hat{L} \in \mathbb{R}\langle \mathbf{x} \rangle_{2t+2}$ of $L$. For this, we set $\hat{L}(p) = L(p)$ for $\deg(p) \leq 2t$, and we set

$$\hat{L}(u^*x_i v) = L(u^*x_i r_v) \quad \text{and} \quad \hat{L}((x_i u)^*x_j v) = L((x_i r_u)^*x_j r_v)$$

for all $i, j \in [n]$ and $u, v \in \langle \mathbf{x} \rangle$ with $|u| = |v| = t$. One can verify that $\hat{L}$ is symmetric and satisfies $x_i(u - r_u) \in N_{t+1}(\hat{L})$ for all $i \in [n]$ and $u \in \mathbb{R}\langle \mathbf{x} \rangle_t$, from which it follows that $\hat{L}$ is 2-flat.

We also have $(u - r_u)x_i \in N_{t+1}(\hat{L})$ for all $i \in [n]$ and $u \in \mathbb{R}\langle \mathbf{x} \rangle_t$: Since $\hat{L}$ is 2-flat, we have $(u - r_u)x_i \in N_{t+1}(\hat{L})$ if and only if $\hat{L}(p(u - r_u)x_i) = 0$ for all $p \in \mathbb{R}\langle \mathbf{x} \rangle_{t-1}$. By using $\deg(x_i p) \leq t$, $L$ is tracial, and $u - r_u \in N_t(L)$, we get $\hat{L}(p(u - r_u)x_i) = L(p(u - r_u)x_i) = L(x_i p(u - r_u)) = 0$.

By consecutively using $(v - r_v)x_j \in N_{t+1}(\hat{L})$, symmetry of $\hat{L}$, $x_i(u - r_u) \in N_{t+1}(\hat{L})$, and again symmetry of $\hat{L}$, we see that

$$\hat{L}((x_i u)^*v x_j) = \hat{L}((x_i u)^*r_v x_j) = \hat{L}((r_v x_j)^*x_i u) = \hat{L}((r_v x_j)^*x_i r_u) = \hat{L}((x_i r_u)^*r_v x_j), \tag{33}$$

and in an analogous way one can show

$$\hat{L}((ux_i)^*x_j v) = \hat{L}((r_u x_i)^*x_j r_v). \tag{34}$$

We can now show that $\hat{L}$ is tracial. We do this by showing that $\hat{L}(wx_j) = \hat{L}(x_j w)$ for all $w$ with $\deg(w) \leq 2t+1$. Notice that when $\deg(w) \leq 2t-1$ the statement follows from the fact that $\hat{L}$ is an extension of $L$. Suppose $w = u^*v$ with $\deg(u) = t+1$ and $\deg(v) \leq t$. We write $u = x_i u'$, and we let $r_{u'}, r_v \in \mathbb{R}\langle \mathbf{x} \rangle_{t-1}$ be such that $u' - r_{u'}, v - r_v \in N_t(L)$. We then have

$$
\begin{aligned}
\hat{L}(wx_j) = \hat{L}(u^*v x_j) &= \hat{L}((x_i u')^*v x_j) \\
&= \hat{L}((x_i r_{u'})^*r_v x_j) && \text{by (33)} \\
&= L((x_i r_{u'})^*r_v x_j) && \text{since } \deg(x_i r_{u'} r_v x_j) \leq 2t \\
&= L((r_{u'} x_j)^*x_i r_v) && \text{since } L \text{ is tracial} \\
&= \hat{L}((r_{u'} x_j)^*x_i r_v) && \text{since } \deg((r_{u'} x_j)^*x_i r_v) \leq 2t \\
&= \hat{L}((u'x_j)^*x_i v) && \text{by (34)} \\
&= \hat{L}(x_j w).
\end{aligned}
$$

It follows $\hat{L}$ is a symmetric tracial flat extension of $L$, and $\operatorname{rank}(M(\hat{L})) = \operatorname{rank}(M(L))$.

Next, we iterate the above procedure to extend $L$ to a symmetric tracial linear functional $\hat{L} \in \mathbb{R}\langle \mathbf{x} \rangle^*$. It remains to show that $\hat{L}$ is nonnegative on $\mathscr{M}(S)$ and zero on $\mathscr{I}(T)$. For this, we make two observations:

(i)  $\mathscr{I}(N_t(L)) \subseteq N(\hat{L})$.
(ii) $\mathbb{R}\langle \mathbf{x} \rangle = \operatorname{span}(W) \oplus \mathscr{I}(N_t(L))$.

For (i) we use the (easy to check) fact that $N_t(L) = \mathrm{span}(\{u - r_u : u \in \langle \mathbf{x} \rangle_t\})$. Then it suffices to show that $w(u - r_u) \in N(\hat{L})$ for all $w \in \langle \mathbf{x} \rangle$, which can be done using induction on $|w|$. From (i), one easily deduces that $\mathrm{span}(W) \cap N(\hat{L}) = \{0\}$, so we have the direct sum $\mathrm{span}(W) \oplus \mathscr{I}(N_t(L))$. The claim (ii) follows using induction on the length of $w \in \langle \mathbf{x} \rangle$: The base case $w \in \langle \mathbf{x} \rangle_t$ follows from (32). Let $w = x_i v \in \langle \mathbf{x} \rangle$ and assume $v \in \mathrm{span}(W) \oplus \mathscr{I}(N_t(L))$, that is, $v = r_v + q_v$ where $r_v \in \mathrm{span}(W)$ and $q_v \in \mathscr{I}(N_t(L))$. We have $x_i v = x_i r_v + x_i q_v$ so it suffices to show $x_i r_v, x_i q_v \in \mathrm{span}(W) \oplus \mathscr{I}(N_t(L))$. Clearly $x_i q_v \in \mathscr{I}(N_t(L))$, since $q_v \in \mathscr{I}(N_t(L))$. Also, observe that $x_i r_v \in \mathbb{R}\langle \mathbf{x} \rangle_t$ and therefore $x_i r_v \in \mathrm{span}(W) \oplus \mathscr{I}(N_t(L))$ by (32).

We conclude the proof by showing that $\hat{L}$ is nonnegative on $\mathscr{M}(S)$ and zero on $\mathscr{I}(T)$. Let $g \in \mathscr{M}(S)$, $h \in \mathscr{I}(T)$, and $p \in \mathbb{R}\langle \mathbf{x} \rangle$. For $p \in \mathbb{R}\langle \mathbf{x} \rangle$ we extend the definition of $r_p$ so that $r_p \in \mathrm{span}(W)$ and $p - r_p \in \mathscr{I}(N_t(L))$, which is possible by (ii). Then,

$$\hat{L}(p^* g p) \overset{\text{(i)}}{=} \hat{L}(p^* g r_p) = \hat{L}(r_p^* g p) \overset{\text{(i)}}{=} \hat{L}(r_p^* g r_p) = L(r_p^* g r_p) \geq 0,$$

$$\hat{L}(p^* h) = \hat{L}(h^* p) \overset{\text{(i)}}{=} \hat{L}(h^* r_p) = \hat{L}(r_p h) = L(r_p h) = 0,$$

where we use $\deg(r_p^* g r_p) \leq 2(t - \delta) + 2\delta = 2t$ and $\deg(r_p h) \leq (t - \delta) + 2\delta \leq 2t$. $\qquad\square$

Combining Theorems 2 and 3 gives the following result, which shows that a flat linear form can be extended to a conic combination of trace evaluation maps. It was first proven in [43, Proposition 6.1] (and in [15] for the unconstrained case).

**Corollary 1** *Let $1 \leq \delta \leq t < \infty$, $S \subseteq \mathrm{Sym}\,\mathbb{R}\langle \mathbf{x} \rangle_{2\delta}$, and $T \in \mathbb{R}\langle \mathbf{x} \rangle_{2\delta}$. If $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ is symmetric, tracial, $\delta$-flat, nonnegative on $\mathscr{M}_{2t}(S)$, and zero on $\mathscr{I}_{2t}(T)$, then it extends to a conic combination of trace evaluations at elements of $\mathscr{D}(S) \cap \mathscr{V}(T)$.*

## A.2 Specialization to the Commutative Setting

The material from Appendix A.1 can be adapted to the commutative setting. Throughout $[\mathbf{x}]$ denotes the set of monomials in $x_1, \ldots, x_n$, i.e., the commutative analog of $\langle \mathbf{x} \rangle$.

The moment matrix $M_t(L)$ of a linear form $L \in \mathbb{R}[\mathbf{x}]_{2t}^*$ is now indexed by the monomials in $[\mathbf{x}]_t$, where we set $M_t(L)_{w,w'} = L(ww')$ for $w, w' \in [\mathbf{x}]_t$. Due to the commutativity of the variables, this matrix is smaller and more entries are now required to be equal. For instance, the $(x_2 x_1, x_3 x_4)$-entry of $M_2(L)$ is equal to its $(x_3 x_1, x_2 x_4)$-entry, which does not hold in general in the noncommutative case.

Given $a \in \mathbb{R}^n$, the *evaluation map* at $a$ is the linear map $L_a \in \mathbb{R}[\mathbf{x}]^*$ defined by

$$L_a(p) = p(a_1, \ldots, a_n) \quad \text{for all} \quad p \in \mathbb{R}[\mathbf{x}].$$

We can view $L_a$ as a trace evaluation at scalar matrices. Moreover, we can view a trace evaluation map at a tuple of pairwise commuting matrices as a conic combination of evaluation maps at scalars by simultaneously diagonalizing the matrices.

The quadratic module $\mathscr{M}(S)$ and the ideal $\mathscr{I}(T)$ have immediate specializations to the commutative setting. We recall that in the commutative setting the (scalar) positivity domain and scalar variety of sets $S, T \subseteq \mathbb{R}[\mathbf{x}]$ are given by[3]

$$D(S) = \left\{ a \in \mathbb{R}^n : g(a) \geq 0 \text{ for } g \in S \right\}, \quad V(T) = \left\{ a \in \mathbb{R}^n : h(a) = 0 \text{ for } h \in T \right\}. \tag{35}$$

We first give the commutative analog of Theorem 1, where we give an additional integral representation in point (3). The equivalence of points (1) and (3) is proved in [64] based on Putinar's Positivstellensatz. Here we give a direct proof on the "moment side" using the Gelfand representation.

**Theorem 4** *Let $S, T \subseteq \mathbb{R}[\mathbf{x}]$ with $\mathscr{M}(S) + \mathscr{I}(T)$ Archimedean. For $L \in \mathbb{R}[\mathbf{x}]^*$, the following are equivalent:*

(1) *$L$ is nonnegative on $\mathscr{M}(S)$, zero on $\mathscr{I}(T)$, and $L(1) = 1$;*
(2) *there exists a unital commutative $C^*$-algebra $\mathscr{A}$ with a state $\tau$ and $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S) \cap \mathscr{V}_{\mathscr{A}}(T)$ such that $L(p) = \tau(p(\mathbf{X}))$ for all $p \in \mathbb{R}[\mathbf{x}]$;*
(3) *there exists a probability measure $\mu$ on $D(S) \cap V(T)$ such that*

$$L(p) = \int_{D(S) \cap V(T)} p(x) \, d\mu(x) \quad \text{for all} \quad p \in \mathbb{R}[\mathbf{x}].$$

***Proof*** $((1) \Rightarrow (2))$ This is the commutative analog of the implication $(1) \Rightarrow (2)$ in Theorem 1 (observing in addition that the operators $X_i$ in (30) pairwise commute so that the constructed $C^*$-algebra $\mathscr{A}$ is commutative).

$((2) \Rightarrow (3))$ Let $\widehat{\mathscr{A}}$ denote the set of unital $*$-homomorphisms $\mathscr{A} \to \mathbb{C}$, known as the *spectrum* of $\mathscr{A}$. We equip $\widehat{\mathscr{A}}$ with the weak-* topology, so that it is compact as a result of $\mathscr{A}$ being unital (see, e.g., [9, II.2.1.4]). The *Gelfand representation* is the $*$-isomorphism

$$\Gamma : \mathscr{A} \to \mathscr{C}(\widehat{\mathscr{A}}), \quad \Gamma(a)(\phi) = \phi(a) \quad \text{for} \quad a \in \mathscr{A}, \ \phi \in \widehat{\mathscr{A}},$$

where $\mathscr{C}(\widehat{\mathscr{A}})$ is the set of complex-valued continuous functions on $\widehat{\mathscr{A}}$. Since $\Gamma$ is an isomorphism, the state $\tau$ on $\mathscr{A}$ induces a state $\tau'$ on $\mathscr{C}(\widehat{\mathscr{A}})$ defined by $\tau'(\Gamma(a)) = \tau(a)$ for $a \in \mathscr{A}$. By the Riesz representation theorem (see, e.g., [67, Theorem 2.14]) there is a Radon measure $\nu$ on $\widehat{\mathscr{A}}$ such that

$$\tau'(\Gamma(a)) = \int_{\widehat{\mathscr{A}}} \Gamma(a)(\phi) \, d\nu(\phi) \quad \text{for all} \quad a \in \mathscr{A}.$$

We then have

$$L(p) = \tau(p(\mathbf{X})) = \tau'(\Gamma(p(\mathbf{X}))) = \int_{\widehat{\mathscr{A}}} \Gamma(p(\mathbf{X}))(\phi) \, d\nu(\phi) = \int_{\widehat{\mathscr{A}}} \phi(p(\mathbf{X})) \, d\nu(\phi)$$

---

[3] Note that in the commutative setting we could avoid using the variety since $V(T) = D(\pm T)$. However, in the noncommutative setting, the polynomials in $T$ need not be symmetric in which case the quadratic module $\mathscr{D}(\pm T)$ would not be well defined.

$$= \int_{\widehat{\mathscr{A}}} p(\phi(X_1), \ldots, \phi(X_n)) \, \mathrm{d}\nu(\phi) = \int_{\widehat{\mathscr{A}}} p(f(\phi)) \, \mathrm{d}\nu(\phi) = \int_{\mathbb{R}^n} p(x) \, \mathrm{d}\mu(x),$$

where $f \colon \widehat{\mathscr{A}} \to \mathbb{R}^n$ is defined by $\phi \mapsto (\phi(X_1), \ldots, \phi(X_n))$, and where $\mu = f_*\nu$ is the pushforward measure of $\nu$ by $f$; that is, $\mu(B) = \nu(f^{-1}(B))$ for measurable $B \subseteq \mathbb{R}^n$.

Since $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S)$, we have $g(\mathbf{X}) \succeq 0$ for all $g \in S$, hence $\Gamma(g(\mathbf{X}))$ is a positive element of $\mathscr{C}(\widehat{\mathscr{A}})$, implying $g(\phi(X_1), \ldots, \phi(X_n)) = \phi(g(\mathbf{X})) = \Gamma(g(\mathbf{X}))(\phi) \geq 0$. Similarly we see $h(\phi(X_1), \ldots, \phi(X_n)) = 0$ for all $h \in T$. So, the range of $f$ is contained in $D(S) \cap V(T)$, $\mu$ is a probability measure on $D(S) \cap V(T)$ since $L(1) = 1$, and we have $L(p) = \int_{D(S) \cap V(T)} p(x) \, \mathrm{d}\mu(x)$ for all $p \in \mathbb{R}[\mathbf{x}]$.

$((3) \Rightarrow (1))$ This is immediate.                                               $\square$

Note that the more common proof for the implication $(1) \Rightarrow (3)$ in Theorem 4 relies on Putinar's Positivstellensatz [64]: if $L$ satisfies (1) then $L(p) \geq 0$ for all polynomials $p$ nonnegative on $D(S) \cap V(T)$ (since $p + \varepsilon \in \mathscr{M}(S) + \mathscr{I}(T)$ for any $\varepsilon > 0$), and thus $L$ has a representing measure $\mu$ as in (3) by the Riesz-Haviland theorem [41].

The following is the commutative analog of Theorem 2.

**Theorem 5** *For $S \subseteq \mathbb{R}[\mathbf{x}]$, $T \subseteq \mathbb{R}[\mathbf{x}]$, and $L \in \mathbb{R}[\mathbf{x}]^*$, the following are equivalent:*

(1) *$L$ is nonnegative on $\mathscr{M}(S)$, zero on $\mathscr{I}(T)$, has $\mathrm{rank}(M(L)) < \infty$, and $L(1) = 1$;*
(2) *there is a finite dimensional commutative $C^*$-algebra $\mathscr{A}$ with a state $\tau$, and $\mathbf{X} \in \mathscr{D}_{\mathscr{A}}(S) \cap \mathscr{V}_{\mathscr{A}}(T)$ such that $L(p) = \tau(p(\mathbf{X}))$ for all $p \in \mathbb{R}[\mathbf{x}]$;*
(3) *$L$ is a convex combination of evaluations at points in $D(S) \cap V(T)$.*

**Proof** $((1) \Rightarrow (2))$ We indicate how to derive this claim from its noncommutative analog. For this denote the commutative version of $p \in \mathbb{R}\langle \mathbf{x} \rangle$ by $p^c \in \mathbb{R}[\mathbf{x}]$. For any $g \in S$ and $h \in T$, select symmetric polynomials $g', h' \in \mathbb{R}\langle \mathbf{x} \rangle$ with $(g')^c = g$ and $(h')^c = h$, and set

$$S' = \left\{ g' : g \in S \right\} \subseteq \mathbb{R}\langle \mathbf{x} \rangle \text{ and } T' = \left\{ h' : h \in T \right\} \cup \left\{ x_i x_j - x_j x_i \right.$$
$$\in \mathbb{R}\langle \mathbf{x} \rangle : i, j \in [n], \, i \neq j \left. \right\} \subseteq \mathbb{R}\langle \mathbf{x} \rangle.$$

Define the linear form $L' \in \mathbb{R}\langle \mathbf{x} \rangle^*$ by $L'(p) = L(p^c)$ for $p \in \mathbb{R}\langle \mathbf{x} \rangle$. Then $L'$ is symmetric, tracial, nonnegative on $\mathscr{M}(S')$, zero on $\mathscr{I}(T')$, and satisfies $\mathrm{rank}\, M(L') = \mathrm{rank}\, M(L) < \infty$. Following the proof of the implication $(1) \Rightarrow (2)$ in Theorem 1, we see that the operators $X_1, \ldots, X_n$ pairwise commute (since $\mathbf{X} \in \mathscr{V}_{\mathscr{A}}(T')$ and $T'$ contains all $x_i x_j - x_j x_i$) and thus the constructed $C^*$-algebra $\mathscr{A}$ is finite dimensional and commutative.

$((2) \Rightarrow (3))$ Here, we follow the proof of this implication in Theorem 2 and observe that since $\mathscr{A}$ is finite dimensional and commutative, it is $*$-isomorphic to an algebra of diagonal matrices ($d_m = 1$ for all $m \in [M]$), which gives directly the desired result.
$((3) \Rightarrow (1))$ is easy.                                                         $\square$

The next result, due to Curto and Fialkow [19], is the commutative analog of Corollary 1.

**Theorem 6** *Let* $1 \leq \delta \leq t < \infty$ *and* $S, T \subseteq \mathbb{R}[\mathbf{x}]_{2\delta}$. *If* $L \in \mathbb{R}[\mathbf{x}]_{2t}^*$ *is* $\delta$-*flat, nonnegative on* $\mathscr{M}_{2t}(S)$, *and zero on* $\mathscr{I}_{2t}(T)$, *then* $L$ *extends to a conic combination of evaluation maps at points in* $D(S) \cap V(T)$.

**Proof** Here too we derive the result from its noncommutative analog in Corollary 1. As in the above proof for the implication (1) $\Longrightarrow$ (2) in Theorem 5, define the sets $S', T' \subseteq \mathbb{R}\langle\mathbf{x}\rangle$ and the linear form $L' \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}^*$ by $L'(p) = L(p^c)$ for $p \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}$. Then $L'$ is symmetric, tracial, nonnegative on $\mathscr{M}_{2t}(S')$, zero on $\mathscr{I}_{2t}(T')$, and $\delta$-flat. By Corollary 1, $L'$ is a conic combination of trace evaluation maps at elements of $\mathscr{D}(S') \cap \mathscr{V}(T')$. It suffices now to observe that such a trace evaluation $L_{\mathbf{X}}$ is a conic combination of (scalar) evaluations at elements of $D(S) \cap V(T)$. Indeed, as $\mathbf{X} \in \mathscr{V}(T')$, the matrices $X_1, \ldots, X_n$ pairwise commute and thus can be assumed to be diagonal. Since $\mathbf{X} \in \mathscr{D}(S') \cap \mathscr{V}(T')$, we have $g(\mathbf{X}) \succeq 0$ for $g' \in S'$ and $h'(\mathbf{X}) = 0$ for $h' \in T'$. This implies $g((X_1)_{jj}, \ldots, (X_n)_{jj}) \geq 0$ and $h((X_1)_{jj}, \ldots, (X_n)_{jj}) = 0$ for all $g \in S$, $h \in T$, and $j \in [d]$. Thus $L_{\mathbf{X}} = \sum_j L_{r_j}$, where $r_j = ((X_1)_{jj}, \ldots, (X_n)_{jj}) \in D(S) \cap V(T)$. $\qquad\square$

Unlike in the noncommutative setting, here we also have the following result, which permits to express any linear functional $L$ nonnegative on an Archimedean quadratic module as a conic combination of evaluations at points, when restricting $L$ to polynomials of bounded degree.

**Theorem 7** *Let* $S, T \subseteq \mathbb{R}[\mathbf{x}]$ *such that* $\mathscr{M}(S) + \mathscr{I}(T)$ *is Archimedean. If* $L \in \mathbb{R}[\mathbf{x}]^*$ *is nonnegative on* $\mathscr{M}(S)$ *and zero on* $\mathscr{I}(T)$, *then for any integer* $k \in \mathbb{N}$ *the restriction of* $L$ *to* $\mathbb{R}[\mathbf{x}]_k$ *extends to a conic combination of evaluations at points in* $D(S) \cap V(T)$.

**Proof** By Theorem 4 there exists a probability measure $\mu$ on $D(S)$ such that

$$L(p) = L(1) \int_{D(S) \cap V(T)} p(x) \, \mathrm{d}\mu(x) \quad \text{for all} \quad p \in \mathbb{R}[\mathbf{x}].$$

A general version of Tchakaloff's theorem, as explained in [4], shows that there exist $r \in \mathbb{N}$, scalars $\lambda_1, \ldots, \lambda_r > 0$ and points $x_1, \ldots, x_r \in D(S)$ such that

$$\int_{D(S) \cap V(T)} p(x) \, \mathrm{d}\mu(x) = \sum_{i=1}^{r} \lambda_i \, p(x_i) \quad \text{for all} \quad p \in \mathbb{R}[\mathbf{x}]_k.$$

Hence the restriction of $L$ to $\mathbb{R}[\mathbf{x}]_k$ extends to a conic combination of evaluations at points in $D(S)$. $\qquad\square$

## A.3 Commutative and Tracial Polynomial Optimization

We briefly discuss here the basic polynomial optimization problems in the commutative and tracial settings. We recall how to design hierarchies of semidefinite programming based bounds, and we give their main convergence properties. The classical commutative polynomial optimization problem asks to minimize a polynomial

$f \in \mathbb{R}[\mathbf{x}]$ over a feasible region of the form $D(S)$ as defined in (35):

$$f_* = \inf_{a \in D(S)} f(a) = \inf\big\{ f(a) : a \in \mathbb{R}^n, \ g(a) \geq 0 \text{ for } g \in S \big\}.$$

In tracial polynomial optimization, given $f \in \operatorname{Sym} \mathbb{R}\langle\mathbf{x}\rangle$, this is modified to minimizing $\operatorname{tr}(f(\mathbf{X}))$ over a feasible region of the form $\mathscr{D}(S)$ as in (6):

$$f_*^{\mathrm{tr}} = \inf_{\mathbf{X} \in \mathscr{D}(S)} \operatorname{tr}(f(\mathbf{X})) = \inf\big\{ \operatorname{tr}(f(\mathbf{X})) : d \in \mathbb{N}, \ \mathbf{X} \in (\mathrm{H}^d)^n, \ g(\mathbf{X}) \succeq 0 \text{ for } g \in S \big\},$$

where the infimum does not change if we replace $\mathrm{H}^d$ by $\mathrm{S}^d$. Commutative polynomial optimization is recovered by restricting to $1 \times 1$ matrices.

For the commutative case, Lasserre [46] and Parrilo [60] have proposed hierarchies of semidefinite programming relaxations based on sums of squares of polynomials and the dual theory of moments. This approach has been extended to eigenvalue optimization [57,61] and later to tracial optimization [14,43]. The starting point in deriving these relaxations is to reformulate the above problems as minimizing $L(f)$ over all normalized trace evaluation maps $L$ at points in $D(S)$ or $\mathscr{D}(S)$, and then to express computationally tractable properties satisfied by such maps $L$.

For $S \cup \{f\} \subseteq \mathbb{R}[\mathbf{x}]$ and $\lceil \deg(f)/2 \rceil \leq t \leq \infty$, recall the (truncated) quadratic module $\mathscr{M}_{2t}(S)$

$$\mathscr{M}_{2t}(S) = \operatorname{cone}\big\{ gp^2 : p \in \mathbb{R}[\mathbf{x}], \ g \in S \cup \{1\}, \ \deg(gp^2) \leq 2t \big\},$$

which we use to formulate the following semidefinite programming lower bound on $f_*$:

$$f_t = \inf\big\{ L(f) : L \in \mathbb{R}[\mathbf{x}]_{2t}^*, \ L(1) = 1, \ L \geq 0 \text{ on } \mathscr{M}_{2t}(S) \big\}.$$

For $t \in \mathbb{N}$ we have $f_t \leq f_\infty \leq f_*$.

In the same way, for $S \cup \{f\} \subseteq \operatorname{Sym} \mathbb{R}\langle\mathbf{x}\rangle$ and $t$ such that $\lceil \deg(f)/2 \rceil \leq t \leq \infty$, we have the following semidefinite programming lower bound on $f_*^{\mathrm{tr}}$:

$$f_t^{\mathrm{tr}} = \inf\big\{ L(f) : L \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}^* \text{ tracial and symmetric}, \ L(1) = 1, \ L \geq 0 \text{ on } \mathscr{M}_{2t}(S) \big\},$$

where we now use definition (1) for $\mathscr{M}_{2t}(S)$.

The next theorem from [46] gives fundamental convergence properties for the commutative case; see also, e.g., [47,49] for a detailed exposition.

**Theorem 8** *Let $1 \leq \delta \leq t < \infty$ and $S \cup \{f\} \subseteq \mathbb{R}[\mathbf{x}]_{2\delta}$ with $D(S) \neq \emptyset$.*

(i) *If $\mathscr{M}(S)$ is Archimedean, then $f_t \to f_\infty$ as $t \to \infty$, the optimal values in $f_\infty$ and $f_*$ are attained, and $f_\infty = f_*$.*

(ii) *If $f_t$ admits an optimal solution $L$ that is $\delta$-flat, then $L$ is a convex combination of evaluation maps at global minimizers of $f$ in $D(S)$, and $f_t = f_\infty = f_*$.*

***Proof*** (i) By repeating the first part of the proof of Theorem 9 in the commutative setting, we see that $f_t \to f_\infty$ and that the optimum is attained in $f_\infty$. Let $L$ be

optimal for $f_\infty$ and let $k$ be greater than $\deg(f)$ and $\deg(g)$ for $g \in S$. By Theorem 7, the restriction of $L$ to $\mathbb{R}[\mathbf{x}]_k$ extends to a conic combination of evaluations at points in $D(S)$. It follows that this extension if feasible for $f_*$ with the same objective value, which shows $f_\infty = f_*$.

(ii) This follows in the same way as the proof of Theorem 9(ii) below, where, instead of using Corollary 1, we now use its commutative analog, Theorem 6.

$\square$

To discuss convergence for the tracial case, we need one more optimization problem:

$$f^{\mathrm{tr}}_{\mathrm{II}_1} = \inf\big\{\tau(f(\mathbf{X})) : \mathbf{X} \in \mathscr{D}_\mathscr{A}(S),\ \mathscr{A} \text{ is a unital } C^*\text{-algebra with tracial state } \tau\big\}.$$

This problem can be seen as an infinite dimensional analog of $f^{\mathrm{tr}}_*$: if we restrict to finite dimensional $C^*$-algebras in the definition of $f^{\mathrm{tr}}_{\mathrm{II}_1}$, then we recover the parameter $f^{\mathrm{tr}}_*$ (use Theorem 2 to see this). Moreover, as we see in Theorem 9(ii) below, equality $f^{\mathrm{tr}}_* = f^{\mathrm{tr}}_{\mathrm{II}_1}$ holds if some flatness condition is satisfied. Whether $f^{\mathrm{tr}}_{\mathrm{II}_1} = f^{\mathrm{tr}}_*$ is true in general is related to Connes' embedding conjecture (see [16,43,44]).

Above we defined the parameter $f^{\mathrm{tr}}_{\mathrm{II}_1}$ using $C^*$-algebras. However, the following lemma shows that we get the same optimal value if we restrict to $\mathscr{A}$ being a von Neumann algebra of type $\mathrm{II}_1$ with separable predual, which is the more common way of defining the parameter $f^{\mathrm{tr}}_{\mathrm{II}_1}$ as is done in [43] (and justifies the notation). We omit the proof of this lemma which relies on a GNS construction and algebraic manipulations, standard for algebraists.

**Lemma 12** *Let $\mathscr{A}$ be a $C^*$-algebra with tracial state $\tau$ and $a_1, \ldots, a_n \in \mathscr{A}$. There exists a von Neumann algebra $\mathscr{F}$ of type $\mathrm{II}_1$ with separable predual, a faithful normal tracial state $\phi$, and elements $b_1, \ldots, b_n \in \mathscr{F}$, so that for every $p \in \mathbb{R}\langle \mathbf{x} \rangle$ we have*

$$\tau(p(a_1, \ldots, a_n)) = \phi(p(b_1, \ldots, b_n)) \quad and$$
$$p(a_1, \ldots, a_n) \text{ is positive} \quad \Longleftrightarrow \quad p(b_1, \ldots, b_n) \text{ is positive}.$$

For all $t \in \mathbb{N}$ we have

$$f^{\mathrm{tr}}_t \le f^{\mathrm{tr}}_\infty \le f^{\mathrm{tr}}_{\mathrm{II}_1} \le f^{\mathrm{tr}}_*,$$

where the last inequality follows by considering for $\mathscr{A}$ the full matrix algebra $\mathbb{C}^{d \times d}$. The next theorem from [43] summarizes convergence properties for these parameters, its proof uses Lemma 13 below.

**Theorem 9** *Let $1 \le \delta \le t < \infty$ and $S \cup \{f\} \subseteq \operatorname{Sym} \mathbb{R}\langle \mathbf{x} \rangle_{2\delta}$ with $\mathscr{D}(S) \ne \emptyset$.*

(i) *If $\mathscr{M}(S)$ is Archimedean, then $f^{tr}_t \to f^{\mathrm{tr}}_\infty$ as $t \to \infty$, and the optimal values in $f^{\mathrm{tr}}_\infty$ and $f^{\mathrm{tr}}_{\mathrm{II}_1}$ are attained and equal.*

(ii) *If $f^{tr}_t$ has an optimal solution $L$ that is $\delta$-flat, then $L$ is a convex combination of normalized trace evaluations at matrix tuples in $\mathscr{D}(S)$, and $f^{tr}_t = f^{tr}_\infty = f^{\mathrm{tr}}_{\mathrm{II}_1} = f^{\mathrm{tr}}_*$.*

@ Springer

**Proof** We first show (i). As $\mathscr{M}(S)$ is Archimedean, $R - \sum_{i=1}^{n} x_i^2 \in \mathscr{M}_{2d}(S)$ for some $R > 0$ and $d \in \mathbb{N}$. Since the bounds $f_t^{\mathrm{tr}}$ are monotone nondecreasing in $t$ and upper bounded by $f_\infty^{\mathrm{tr}}$, the limit $\lim_{t \to \infty} f_t^{\mathrm{tr}}$ exists and it is at most $f_\infty^{\mathrm{tr}}$.

Fix $\varepsilon > 0$. For $t \in \mathbb{N}$ let $L_t$ be a feasible solution to the program defining $f_t^{\mathrm{tr}}$ with value $L_t(f) \leq f_t^{\mathrm{tr}} + \varepsilon$. As $L_t(1) = 1$ for all $t$ we can apply Lemma 13 below and conclude that the sequence $(L_t)_t$ has a convergent subsequence. Let $L \in \mathbb{R}\langle \mathbf{x} \rangle^*$ be the pointwise limit. One can easily check that $L$ is feasible for $f_\infty^{\mathrm{tr}}$. Hence we have $f_\infty^{\mathrm{tr}} \leq L(f) \leq \lim_{t \to \infty} f_t^{\mathrm{tr}} + \varepsilon \leq f_\infty^{\mathrm{tr}} + \varepsilon$. Letting $\varepsilon \to 0$ we obtain that $f_\infty^{\mathrm{tr}} = \lim_{t \to \infty} f_t^{\mathrm{tr}}$ and $L$ is optimal for $f_\infty^{\mathrm{tr}}$.

Next, since $L$ is symmetric, tracial, and nonnegative on $\mathscr{M}(S)$, we can apply Theorem 1 to obtain a feasible solution $(\mathscr{A}, \tau, \mathbf{X})$ to $f_{\mathrm{II}_1}^{\mathrm{tr}}$ satisfying (29) with objective value $L(f)$. This shows $f_\infty^{\mathrm{tr}} = f_{\mathrm{II}_1}^{\mathrm{tr}}$ and that the optima are attained in $f_\infty^{\mathrm{tr}}$ and $f_{\mathrm{II}_1}^{\mathrm{tr}}$.

Finally, part (ii) is derived as follows. If $L$ is an optimal solution of $f_t^{\mathrm{tr}}$ that is $\delta$-flat, then, by Corollary 1, it has an extension $\hat{L} \in \mathbb{R}\langle \mathbf{x} \rangle^*$ that is a conic combination of trace evaluations at elements of $\mathscr{D}(S)$. This shows $f_*^{\mathrm{tr}} \leq \hat{L}(f) = L(f)$, and thus the chain of equalities $f_t^{\mathrm{tr}} = f_\infty^{\mathrm{tr}} = f_*^{\mathrm{tr}} = f_{\mathit{\Pi}_1}^{\mathrm{tr}}$ holds.                                    $\square$

We conclude with the following technical lemma, based on the Banach-Alaoglu theorem. It is a well-known crucial tool for proving the asymptotic convergence result from Theorem 9(i) and it is used at other places in the paper.

**Lemma 13** *Let $S \subseteq \mathrm{Sym}\,\mathbb{R}\langle \mathbf{x} \rangle$, $T \subseteq \mathbb{R}\langle \mathbf{x} \rangle$, and assume $R - (x_1^2 + \cdots + x_n^2) \in \mathscr{M}_{2d}(S) + \mathscr{I}_{2d}(T)$ for some $d \in \mathbb{N}$ and $R > 0$. For $t \in \mathbb{N}$ assume $L_t \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ is tracial, nonnegative on $\mathscr{M}_{2t}(S)$ and zero on $\mathscr{I}_{2t}(T)$. Then we have $|L_t(w)| \leq R^{|w|/2} L_t(1)$ for all $w \in \langle \mathbf{x} \rangle_{2t-2d+2}$. In addition, if $\sup_t L_t(1) < \infty$, then $\{L_t\}_t$ has a pointwise converging subsequence in $\mathbb{R}\langle \mathbf{x} \rangle^*$.*

**Proof** We first use induction on $|w|$ to show that $L_t(w^*w) \leq R^{|w|} L_t(1)$ for all $w \in \langle \mathbf{x} \rangle_{t-d+1}$. For this, assume $L_t(w^*w) \leq R^{|w|} L_t(1)$ and $|w| \leq t - d$. Then we have

$$L_t((x_i w)^* x_i w) = L_t(w^*(x_i^2 - R)w) + R \cdot L_t(w^*w) \leq R \cdot R^{|w|} L_t(1) = R^{|x_i w|} L_t(1).$$

For the inequality we use the fact that $L_t(w^*(x_i^2 - R)w) \leq 0$ since $w^*(R - x_i^2)w$ can be written as the sum of a polynomial in $\mathscr{M}_{2t}(S) + \mathscr{I}_{2t}(T)$ and a sum of commutators of degree at most $2t$, which follows using the following identity: $w^*qhw = ww^*qh + [w^*qh, w]$. Next we write any $w \in \langle \mathbf{x} \rangle_{2(t-d+1)}$ as $w = w_1^* w_2$ with $w_1, w_2 \in \langle \mathbf{x} \rangle_{t-d+1}$ and use the positive semidefiniteness of the principal submatrix of $M_t(L_t)$ indexed by $\{w_1, w_2\}$ to get

$$L_t(w)^2 = L_t(w_1^* w_2)^2 \leq L_t(w_1^* w_1) L_t(w_2^* w_2) \leq R^{|w_1| + |w_2|} L_t(1)^2 = R^{|w|} L_t(1)^2.$$

This shows the first claim.

Suppose $c := \sup_t L_t(1) < \infty$. For each $t \in \mathbb{N}$, consider the linear functional $\hat{L}_t \in \mathbb{R}\langle \mathbf{x} \rangle^*$ defined by $\hat{L}_t(w) = L_t(w)$ if $|w| \leq 2t - 2d + 2$ and $\hat{L}_t(w) = 0$ otherwise. Then the vector $(\hat{L}_t(w)/(cR^{|w|/2}))_{w \in \langle \mathbf{x} \rangle}$ lies in the supremum norm unit ball of $\mathbb{R}^{\langle \mathbf{x} \rangle}$, which is compact in the weak$*$ topology by the Banach–Alaoglu theorem.

It follows that the sequence $(\hat{L}_t)_t$ has a pointwise converging subsequence and thus the same holds for the sequence $(L_t)_t$. □

## References

1. M.F. Anjos and J.B. Lasserre. Handbook on Semidefinite, Conic and Polynomial Optimization. International Series in Operations Research & Management Science Series, Springer, 2012.
2. A. Atserias, L. Mančinska, D. Roberson, R. Šámal, S. Severini, and A. Varvitsiotis. Quantum and non-signalling graph isomorphisms. *Journal of Combinatorial Theory, Series B* (2018). https://doi.org/10.1016/j.jctb.2018.11.002.
3. G.P. Barker, L.Q. Eifler, and T.P. Kezlan. A non-commutative spectral theorem, *Linear Algebra and its Applications* 20(2) (1978), 95–100.
4. C. Bayer, J. Teichmann. The proof of Tchakaloff's theorem. *Proceedings of the American Mathematical Society* 134 (2006), 3035–3040.
5. A. Berman, U.G. Rothblum. A note on the computation of the cp-rank. *Linear Algebra and its Applications* 419 (2006), 1–7.
6. A. Berman, N. Shaked-Monderer. Completely Positive Matrices. World Scientific, 2003.
7. M. Berta, O. Fawzi, V.B. Scholz. Quantum bilinear optimization. *SIAM Journal on Optimization* 26(3) (2016), 1529–1564.
8. J. Bezanson, A. Edelman, S. Karpinski, V.B. Shah. Julia: A Fresh Approach to Numerical Computing. *SIAM Review* 59(1) (2017), 65–98.
9. B. Blackadar. Operator Algebras: Theory of C*-Algebras and Von Neumann Algebras. Encyclopaedia of Mathematical Sciences, Springer, 2006.
10. I.M. Bomze, W. Schachinger, R. Ullrich. From seven to eleven: Completely positive matrices with high cp-rank. *Linear Algebra and its Applications* 459 (2014), 208 – 221.
11. I.M. Bomze, W. Schachinger, R. Ullrich. New lower bounds and asymptotics for the cp-rank. *SIAM Journal on Matrix Analysis and Applications* 36 (2015), 20–37.
12. G. Braun, S. Fiorini, S. Pokutta, D. Steurer. Approximation limits of linear programs (beyond hierarchies). *Mathematics of Operations Research* 40(3) (2015), 756–772. Appeared earlier in FOCS'12.
13. S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming* 120(2) (2009), 479–495.
14. S. Burgdorf, K. Cafuta, I. Klep, J. Povh. The tracial moment problem and trace-optimization of polynomials. *Mathematical Programming* 137(1) (2013), 557–578.
15. S. Burgdorf, I. Klep. The truncated tracial moment problem. *Journal of Operator Theory* 68(1) (2012), 141–163.
16. S. Burgdorf, I. Klep, J. Povh. Optimization of Polynomials in Non-Commutative Variables. Springer Briefs in Mathematics, Springer, 2016.
17. S. Burgdorf, M. Laurent, T. Piovesan. On the closure of the completely positive semidefinite cone and linear approximations to quantum colorings. *Electronic Journal of Linear Algebra* 32 (2017), 15–40.
18. M. Conforti, G. Cornuéjols, G. Zambelli. Extended formulations in combinatorial optimization. *4OR* 8 (2010), 1–48.
19. R.E. Curto, L.A. Fialkow. Solution of the Truncated Complex Moment Problem for Flat Data. Memoirs of the American Mathematical Society, American Mathematical Society, 1996.
20. P. Dickinson, M. Dür. Linear-time complete positivity detection and decomposition of sparse matrices. *SIAM Journal on Matrix Analysis and Applications* 33(3) (2012), 701–720.
21. J.H. Drew, C.R. Johnson, R. Loewy. Completely positive matrices associated with M-matrices. *Linear and Multilinear Algebra* 37(4) (1994), 303–310.
22. K.J. Dykema, V.I. Paulsen, J. Prakash. Non-closure of the set of quantum correlations via graphs, arXiv:1709.05032 (2017).
23. J. Edmonds. Maximum matching and a polyhedron with 0, 1 vertices. *Journal of Research of the National Bureau of Standards* 69 B (1965), 125–130.
24. Y. Faenza, S. Fiorini, R. Grappe, H. Tiwari. Extended formulations, non-negative factorizations and randomized communication protocols. *Mathematical Programming* 153(1) (2015), 75–94.
25. H. Fawzi, J. Gouveia, P.A. Parrilo, R.Z. Robinson, R.R. Thomas. Positive semidefinite rank. *Mathematical Programming* 153(1) (2015), 133–177.

26. H. Fawzi, P.A. Parrilo. Lower bounds on nonnegative rank via nonnegative nuclear norms. *Mathematical Programming* 153(1) (2015), 41–66.
27. H. Fawzi, P.A. Parrilo. Self-scaled bounds for atomic cone ranks: applications to nonnegative rank and cp-rank. *Mathematical Programming* 158(1) (2016), 417–465.
28. S. Fiorini, V. Kaibel, K. Pashkovich, D. Theis. Combinatorial bounds on nonnegative rank and extended formulations. *Discrete Mathematics* 313(1) (2013), 67–83.
29. S. Fiorini, S. Massar, S. Pokutta, H.R. Tiwary, R. de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *Journal of the ACM* 62(2) (2015), 17:1–17:23. Appeared earlier in STOC'12.
30. N. Gillis. Introduction to nonnegative matrix factorization. *SIAG/OPT Views and News* 25(1) (2017), 7–16.
31. N. Gillis, F. Glineur. On the geometric interpretation of the nonnegative rank. *Linear Algebra and its Applications* 437(11) (2012), 2685–2712.
32. M. Goemans. Smallest compact formulation for the permutahedron. *Mathematical Programming* 153(1) (2015), 5–11.
33. A.P. Goucha, J. Gouveia, P.M. Silva. On ranks of regular polygons. *SIAM Journal on Discrete Mathematics* 31(4) (2016), 2612–2625.
34. J. Gouveia, P.A. Parrilo, R.R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research* 38(2) (2013), 248–264.
35. J. Gouveia, R.Z. Robinson, R.R. Thomas. Polytopes of minimum positive semidefinite rank. *Discrete & Computational Geometry* 50(3) (2013), 679–699.
36. M. Grant, S. Boyd. CVX: Matlab Software for Disciplined Convex Programming, version 2.1, 2014. http://cvxr.com/cvx
37. S. Gribling, D. de Laat, M. Laurent. Bounds on entanglement dimensions and quantum graph parameters via noncommutative polynomial optimization. *Mathematical Programming Series B* 171(1) (2018), 5–42.
38. S. Gribling, D. de Laat, M. Laurent. Matrices with high completely positive semidefinite rank. *Linear Algebra and its Applications* 513 (2017), 122 – 148.
39. P. Groetzner, M. Dür. A factorization method for completely positive matrices. Preprint (2018), http://www.optimization-online.org/DB_HTML/2018/03/6511.html.
40. M. Grötschel, L. Lovász., A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica* 1(2) (1981), 169–197.
41. E.K. Haviland. On the Momentum Problem for Distribution Functions in More Than One Dimension. II. *American Journal of Mathematics* 58(1) (1936), 164–168.
42. R. Jain, Y. Shi, Z. Wei, S. Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. *IEEE Transactions on Information Theory* 59(8) (2013), 5171–5178.
43. I. Klep, J. Povh. Constrained trace-optimization of polynomials in freely noncommuting variables. *Journal of Global Optimization* 64(2) (2016), 325–348.
44. I. Klep, M. Schweighofer. Connes' embedding conjecture and sums of hermitian squares. *Advances in Mathematics* 217(4) (2008), 1816–1837.
45. E. de Klerk, D.V. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM Journal on Optimization* 12(4) (2002), 875–892.
46. J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization* 11(3) (2001), 796–817.
47. J.B. Lasserre. Moments, Positive Polynomials and Their Applications, . Imperial College Press, 2009.
48. J.B. Lasserre. New approximations for the cone of copositive matrices and its dual. *Mathematical Programming* 144(1-2) (2014), 265–276.
49. M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In Emerging Applications of Algebraic Geometry (M. Putinar, S. Sullivant eds.), Springer, 2009, pp. 157–270.
50. M. Laurent, T. Piovesan. Conic approach to quantum graph parameters using linear optimization over the completely positive semidefinite cone. *SIAM Journal on Optimization* 25(4) (2015), 2461–2493.
51. J.R. Lee, P. Raghavendra, D. Steurer. Lower bounds on the size of semidefinite programming relaxations. In Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing, STOC'15, 2015, pp. 567–576.
52. T. Lee, Z. Wei, R. de Wolf. Some upper and lower bounds on psd-rank. *Mathematical Programming* 162(1) (2017), 495–521.

53. L. Mančinska, D. Roberson. Note on the correspondence between quantum correlations and the completely positive semidefinite cone. Available at quantu-minfo.quantumlah.org/memberpages/laura/corr.pdf (2014).

54. R.K. Martin. Using separation algorithms to generate mixed integer model reformulations. *Operations Research Letters* 10(3) (1991), 119–128.

55. D. Mond, J. Smith, D. van Straten. Stochastic factorizations, sandwiched simplices and the topology of the space of explanations. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **459**(2039) (2003), 2821–2845.

56. MOSEK ApS. The MOSEK optimization toolbox for MATLAB manual. Version 8.0.0.81, 2017. URL http://docs.mosek.com/8.0/toolbox.pdf

57. M. Navascués, S. Pironio, A. Acín. SDP relaxations for non-commutative polynomial optimization. In Handbook on Semidefinite, Conic and Polynomial Optimization (M.F. Anjos, J.B. Lasserre eds.). Springer, 2012, pp. 601–634.

58. J. Nie. The $\mathscr{A}$-truncated $K$-moment problem. *Foundations of Computational Mathematics* 14(6) (2014), 1243–1276.

59. J. Nie. Symmetric tensor nuclear norms. *SIAM Journal on Applied Algebra and Geometry* 1(1) (2017), 599–625.

60. P.A. Parrilo. Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization. PhD thesis, Caltech, 2000.

61. S. Pironio, M. Navascués, A. Acín. Convergent relaxations of polynomial optimization problems with noncommuting variables. *SIAM Journal on Optimization* 20(5) (2010), 2157–2180.

62. A. Prakash, J. Sikora, A. Varvitsiotis, Z. Wei. Completely positive semidefinite rank. *Mathematical Programming* 171(1–2) (2017), 397–431.

63. A. Prakash, A. Varvitsiotis. Correlation matrices, Clifford algebras, and completely positive semidefinite rank. *Linear and Multilinear Algebra* (2018). https://doi.org/10.1080/03081087.2018.1529136.

64. Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana University Mathematics Journal **42**, 969–984 (1993)

65. J. Renegar. On the computational complexity and geometry of the first-order theory of the reals. Part I: Introduction. Preliminaries. The geometry of semi-algebraic sets. The decision problem for the existential theory of the reals. *Journal of Symbolic Computation* 13(3) (1992), 255 – 299.

66. T. Rothvoss. The matching polytope has exponential extension complexity. In Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing, STOC'14, 2014, pp. 263–272.

67. W. Rudin. Real and complex analysis. Mathematics series. McGraw-Hill, 1987.

68. N. Shaked-Monderer, A. Berman, I.M. Bomze, F. Jarre, W. Schachinger. New results on the cp-rank and related properties of co(mpletely )positive matrices. *Linear and Multilinear Algebra* 63(2) (2015), 384–396.

69. N. Shaked-Monderer, I.M. Bomze, F. Jarre, W. Schachinger. On the cp-rank and minimal cp factorizations of a completely positive matrix. *SIAM Journal on Matrix Analysis and Applications* 34(2) (2013), 355–368.

70. Y. Shitov. A universality theorem for nonnegative matrix factorizations. arXiv:1606.09068v2 (2016).

71. Y. Shitov. The complexity of positive semidefinite matrix factorization. *SIAM Journal on Optimization* 27(3) (2017), 1898–1909.

72. J. Sikora, A. Varvitsiotis. Linear conic formulations for two-party correlations and values of nonlocal games. *Mathematical Programming* 162(1) (2017), 431–463.

73. W. Slofstra. The set of quantum correlations is not closed. arXiv:1703.08618 (2017).

74. G. Tang, P. Shah. Guaranteed tensor decomposition: A moment approach. In Proceedings of the 32nd International Conference on International Conference on Machine Learning, ICML'15, 2015, pp. 1491–1500.

75. A. Vandaele, F. Glineur, N. Gillis. Algorithms for positive semidefinite factorization. *Computational Optimization and Applications* 71(1) (2018), 193–219.

76. S.A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization* 20(3) (2009), 1364–1377.

77.  J.H.M. Wedderburn. Lectures on Matrices. Dover Publications Inc., 1964.

78.  M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences* 43(3) (1991), 441 – 466.

**Publisher's Note**  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.