Lower bounds on matrix factorization ranks
via noncommutative polynomial optimization

Sander Gribling[*]   David de Laat[†]   Monique Laurent[‡]

August 7, 2017

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 completely positive rank, and their symmetric analogues: the positive semidefinite rank and the completely positive semidefinite rank. We study the convergence properties of our hierarchies, compare them extensively to known lower bounds, and provide some (numerical) examples.

Contents

1

1 Introduction

1.1 Matrix factorization ranks

A factorization of a matrix $A \in \mathbb{R}^{m \times n}$ over a sequence of cones $\{K^d\}$, each equipped with an inner product, is a decomposition of the form $A_{ij} = \langle X_i, Y_j \rangle$ with $X_i, Y_j \in K^d$ for all i, j. The smallest d for which such a factorization exists is called the *cone factorization rank* of A over $\{K^d\}$ [GPT13]. An important example is the *nonnegative rank*, denoted $\operatorname{rank}_+(A)$, where K^d is the nonnegative orthant $\mathbb{R}_+^d$ with the standard inner product. Another important example is the *positive semidefinite rank*, denoted $\operatorname{psd-rank}_{\mathbb{R}}(A)$, where K^d is the positive semidefinite cone $\mathcal{S}_+^d$ with the *trace inner product* $\langle X, Y \rangle = \operatorname{Tr}(X^\mathsf{T} Y)$. Both the nonnegative rank and 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 [Yan91], 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 is the linear image of an affine section of the nonnegative orthant $\mathbb{R}_+^d$, and the slack matrix of P is given by $(b_i - a_i^\mathsf{T} v)_{v \in V, i \in I}$ when $P = \operatorname{conv}(V) = \{x : a_i^\mathsf{T} x \le b_i \ (i \in I)\}$. 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 $\mathcal{S}_+^d$ [GPT13]. The motivation to study extension complexity is that polytopes with small extension complexity admit efficient algorithms for linear optimization. Well-known examples include spanning tree polytopes [Mar91] and permutahedra [Goe15], which have polynomial linear extension complexity, and the stable set polytope of perfect graphs, which has polynomial semidefinite extension complexity [MGS81] (see, e.g., the surveys [CCZ10, FGP+15]). The above connection to matrix factorization ranks 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 [FMP+15], and this result was extended to their semidefinite extension complexities in [LRS15]. Surprisingly, the linear extension complexity of the matching polytope is also exponential [Rot14], even though linear optimization over this set is polynomial time solvable [Edm65]. It is an open question whether the semidefinite extension complexity of the matching polytope is exponential.

Besides applications in extension complexity, the nonnegative rank also finds applications in probability and communication complexity, and the positive semidefinite rank has applications in quantum information theory and quantum communication complexity (see, e.g., [MSvS03, FFGT15, JSWZ13, FMP+15]).

We are also interested in *symmetric* analogues of the above matrix factorization ranks, where we require the same factors for the rows and columns ($m = n$ and $X_i = Y_i$ for all i). Analogously to the nonnegative rank, we consider the *completely positive rank*, denoted cp-rank(A), where $K^d = \mathbb{R}_+^d$. We also consider the *completely positive semidefinite rank*, denoted cpsd-rank$_{\mathbb{K}}(A)$, with $K^d = \mathcal{S}_+^d$ if $\mathbb{K} = \mathbb{R}$ and $K^d = \mathcal{H}_+^d$ if $\mathbb{K} = \mathbb{C}$, where $\mathcal{H}_+^d$ is the cone of complex Hermitian positive semidefinite matrices with the trace inner product $\langle X, Y \rangle = \operatorname{Tr}(X^* Y)$. The symmetric matrices for which these parameters are well defined form convex cones known as the *completely positive cone*, denoted $\mathcal{CP}^n$, and the *completely positive semidefinite cone*, denoted $\mathcal{CS}_+^n$. We have the inclusions $\mathcal{CP}^n \subseteq \mathcal{CS}_+^n \subseteq \mathcal{S}_+^n$, which are strict for $n \ge 5$. For details on these cones see [BSM03, LP15].

Motivation for the cones $\mathcal{CP}^n$ and $\mathcal{CS}_+^n$ comes in particular from their use to model classical and quantum information optimization problems. Graph parameters such as the stability number and the chromatic number can be written as linear optimization problems over the completely positive cone [dKP02] and the same holds, more generally, for any quadratic problem with mixed binary variables [Bur09]. The cp-rank is widely studied in the linear algebra community; see, e.g., [BSM03, SMBJS13, SMBB+15, BSU14]. The completely positive semidefinite cone was first studied in [LP15] to describe quantum analogues of the stability number and chro-

matic number of a graph, which was later extended to general graph homomorphisms [SV17] and graph isomorphism [AMR+16]. An additional connection between the completely positive semidefinite cone and the set of quantum correlations is shown in [MR14, SV17]. 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 [PSVW16, GdLL17a, PV17] 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 given by the completely positive semidefinite rank of a certain matrix (see [GdLL17b]).

We may also consider the parameter psd-rank$_{\mathbb{C}}(A)$, obtained by using asymmetric factorizations by complex Hermitian positive semidefinite matrices. We have psd-rank$_{\mathbb{C}}(A) \leq$ psd-rank$_{\mathbb{R}}(A) \leq \text{rank}_+(A) \leq \min\{m, n\}$ and cp-rank$(A) \leq \binom{n+1}{2}$, but 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 [Slo17] that the cone $\mathcal{CS}_+^n$ is not closed for $n \geq 1942$, so 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 $\mathcal{CS}_+^5$ is closed, and we do not know how to construct a 5×5 matrix whose cpsd-rank exceeds 5.

The rank$_+$, cp-rank, and psd-rank are known to be computable (using results from [Ren92], since upper bounds exist on these factorization ranks; see [BR06] for cp-rank). But computing the nonnegative rank is NP-complete [Vav09] and the psd-rank is existential theory of the reals complete [Shi16]. For the cp-rank and 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 one can employ heuristics that try to construct small factorizations. Many such heuristics exist for the nonnegative rank (see the overview [Gil17] and references therein), factorization algorithms exist for structured completely positive matrices (see [DD12]), and algorithms to compute positive semidefinite factorizations are presented in the recent work [VGG17]. In this paper we want to compute *lower* bounds, 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 above mentioned matrix factorization ranks, based on tools from (noncommutative) polynomial optimization.

We introduce in Section 3 our approach for the completely positive semidefinite rank. We start by defining a hierarchy of lower bounds

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

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 $\sqrt{\text{rank}(A)} \leq \text{cpsd-rank}_{\mathbb{C}}(A)$ follows by viewing a Hermitian $d \times d$ matrix as a d^2-dimensional real vector, and an analytic lower bound is given in [PSVW16]. We show $\xi_1^{\text{cpsd}}(A)$ 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 not exist. We also introduce several ideas leading to strengthenings of the basic bounds $\xi_t^{\text{cpsd}}(A)$. 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 completely positive rank much more is known about lower bounds. The best known generic lower bounds are due to Fawzi and Parrilo [FP15, FP16].

In [FP16] 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 [FP16] it is also shown that $\tau_+(A)$ is at least as good as certain norm-based lower bounds. In particular, τ_+ is at least as good as the ℓ_∞ norm-based lower bound, which was used by Rothvoß [Rot14] to show that the matching polytope has exponential linear extension complexity. In [FP15] 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 τ_+.

Fawzi and Parrilo [FP16] 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) is equal to the smallest d for which A can be written as $A = \sum_{k=1}^{d} R_k$, where each R_k is a nonnegative (positive semidefinite) rank one matrix. As the psd-rank and cpsd-rank are not known to admit atomic formulations, the techniques from [FP16] do not extend directly to these factorization ranks. However, our approach via polynomial optimization permits to capture these factorization ranks as well.

In Sections 4 and 5 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 $\xi_t^+(A)$ converges to $\tau_+(A)$ as $t \to \infty$. The basic hierarchy $\xi_t^{\mathrm{cp}}(A)$ for the cp-rank does not converge to $\tau_{\mathrm{cp}}(A)$, but we provide two types of additional constraints that can be added to $\xi_t^{\mathrm{cp}}(A)$ 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 is at least $\tau_{\mathrm{cp}}(A) - \varepsilon$. We give numerical results on small matrices studied in the literature, and we show that $\xi_3^+(A)$ can improve over $\tau_+^{\mathrm{sos}}(A)$ for examples studied in [FP16].

Finally, in Section 6 we derive a hierarchy $\xi_t^{\mathrm{psd}}(A)$ of lower bounds on the psd-rank. We compare the bounds $\xi_t^{\mathrm{psd}}(A)$ to a bound from [LWdW17] and we provide some numerical examples illustrating the performance of our bounds.

We provide two implementations of all lower bounds introduced in this paper, at the arXiv submission of this paper. One implementation uses Matlab and the CVX package [GB14], and the other uses Julia [BEKS17]. The implementations support various SDP solvers, for our numerical examples we used Mosek [ApS17].

1.3 Basic approach and notation

We use (noncommutative) polynomial optimization to define hierarchies of relaxations for the problem of finding smallest possible matrix factorization ranks.

In classical polynomial optimization the problem is to find the global minimum of a polynomial over a semialgebraic set of the form $\{x \in \mathbb{R}^n : g(x) \geq 0 \text{ for } g \in S\}$, where $S \subseteq \mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \ldots, x_n]$. Lasserre [Las01] and Parrilo [Par00] have proposed hierarchies of semidefinite programming relaxations based on the theory of moments and the dual theory of sums of squares polynomials, which can be used to compute successively better lower bounds converging to the global minimum (under an Archimedean condition). This approach has been used in a wide range of applications and there is extensive literature (see, e.g., [AL12, Las09, Lau09]). Most relevant to this work, it is used in [Las14] to design conic approximations of the completely positive cone and in [Nie14] to check membership in the completely positive cone.

This approach has also been extended to eigenvalue optimization [PNA10, NPA12], and later to tracial optimization [BCKP13, KP16]. Here we briefly explain the tracial optimization problem, which is most relevant to our work, for which we first need 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 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. We can evaluate a noncommutative polynomial at a matrix tuple $\mathbf{X} = (X_1, \ldots, X_n) \in (\mathcal{H}^d)^n$. In tracial optimization the problem is to minimize the normalized trace $\mathrm{tr}(f(\mathbf{X}))$ over the matrix positivity domain

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

where $S \subseteq \mathrm{Sym}\,\mathbb{R}\langle\mathbf{x}\rangle$. Throughout $\mathrm{tr}(\cdot)$ denotes the *normalized trace*, while $\mathrm{Tr}(\cdot)$ denotes the (non-normalized) trace, so that $\mathrm{tr}(I) = 1$ and $\mathrm{Tr}(I) = d$ for the identity matrix $I \in \mathcal{H}^d$. As in the classical (commutative) case, semidefinite programming hierarchies have been constructed to obtain lower bounds on the minimal normalized trace. Here the distinguishing feature is the dimension independence: the optimization is over all possible matrix sizes. Perhaps counterintuitively, in this paper we use these techniques to compute lower bounds on factorization dimensions.

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

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

Then $A = (L_{\mathbf{X}}(x_i x_j))$ and $\mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A) = d = L(1)$. 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 been used recently [TS15, Nie16] in the commutative setting to derive a converging hierarchy to the nuclear norm of a symmetric tensor. In order to explain how we can give a semidefinite programming hierarchy using this idea we first need some more notation.

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. Let $\mathbb{R}\langle\mathbf{x}\rangle_t^*$ be the space of real-valued linear functionals on $\mathbb{R}\langle\mathbf{x}\rangle_t$. 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 $p \in \mathbb{R}\langle\mathbf{x}\rangle_t$. Given 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 $2t$, denoted by $\mathcal{M}_{2t}(S)$, is the cone generated by all polynomials $p^*gp \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}$ with $g \in S \cup \{1\}$:

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

The linear functional $L_{\mathbf{X}}$ as defined above 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 $L_{\mathbf{X}}(p^*p) \geq 0$ and, as will be explained later, $L_{\mathbf{X}}$ satisfies the following *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:

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

The bound $\xi_t^{\mathrm{cpsd}}(A)$ is computationally tractable (for small t). Indeed the localizing constraint ($L \geq 0$ on $\mathcal{M}_{2t}(S)$) 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. These localizing conditions ensure the Archimedean property for the quadratic module, which permits to show some convergence properties of the bounds $\xi_t^{\mathrm{cpsd}}(A)$.

The above idea extends naturally for 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.

1.4 Organization

We start by providing the necessary background about commutative and tracial polynomial optimization in Section 2, with proofs relegated to the appendix. We then introduce in Section 3 the basic hierarchy of lower bounds on the cpsd-rank, including a discussion of its properties and of some ways to improve the 'basic' bounds. In Sections 4 and 5 we present the adapted hierarchies for the completely positive rank and the nonnegative rank, respectively. As mentioned above, we compare our results to known bounds in [FP16], and we also provide computational examples illustrating the behaviour of the various bounds. In Section 6 we discuss our hierarchy for the positive semidefinite rank, and we also provide some computational examples.

2 Commutative and tracial polynomial optimization

In this section we discuss known convergence and flatness results for commutative and tracial optimization. We present these results in such a way that we can later use them for our hierarchies to lower bound matrix factorization ranks. Although the commutative case was developed first, here we treat the commutative and tracial cases together. We will provide proofs (in the appendices) working only 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 [PNA10], but here we only discuss the commutative and tracial cases, as these are the most relevant to our approach.

2.1 Flat extensions and representations of linear forms

We start by defining what it means for a linear functional $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ to be flat. For this we can use the null space of L, which is 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 also use the notation $N(L) = N_\infty(L)$. Given an integer $1 \leq \delta \leq t$, the functional L is said to be δ-*flat* if

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

Moreover, we say that L is *flat* if L is δ-flat for some $\delta \geq 1$.

Many properties satisfied by a linear functional $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ can be conveniently expressed in terms of its *moment matrix* (aka *Hankel matrix*) $M_t(L)$, which is the matrix whose rows and columns are indexed by the words in $\langle \mathbf{x} \rangle_t$, with entries

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

Indeed, the linear form L is symmetric if and only if $M_t(L)$ is symmetric, and L is symmetric and positive if and only if $M_t(L)$ is positive semidefinite. Moreover, the null space of L can 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 $M_t(L)$. Finally, the form L is δ-flat if and only if the rank of $M_t(L)$ is equal to the rank of its principal submatrix indexed by the words in $\langle \mathbf{x} \rangle_{t-\delta}$; that is, if

$$\operatorname{rank} M_t(L) = \operatorname{rank} M_{t-\delta}(L).$$

One can express nonnegativity of a tracial linear form $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ on a truncated quadratic module $\mathcal{M}_{2t}(S)$ as defined in (1) in terms of certain associated positive semidefinite moment matrices: Given a polynomial $g \in \mathbb{R}\langle \mathbf{x} \rangle$, define $gL \in \mathbb{R}\langle \mathbf{x} \rangle_{2t-\deg(g)}^*$ by $(gL)(p) = L(gp)$, so that

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

and thus $L \geq 0$ on $\mathcal{M}_{2t}(S)$ if and only if $M_{t-d_g}(gL) \succeq 0$ for all $g \in S \cup \{1\}$.

Important examples of positive symmetric tracial linear forms on $\mathbb{R}\langle \mathbf{x} \rangle$ arise from trace evaluation maps. Given a tuple $\mathbf{X} = (X_1, \ldots, X_n) \in (\mathcal{H}^d)^n$ of Hermitian $d \times d$ matrices, we define the *trace evaluation map* $L_{\mathbf{X}}$ on $\mathbb{R}\langle \mathbf{x} \rangle$ by setting

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

Recall $\operatorname{Tr}(\cdot)$ is the matrix trace that satisfies $\operatorname{Tr}(I) = d$ for $I \in \mathcal{H}^d$. If we instead use the *normalized trace* $\operatorname{tr}(\cdot)$, then $L_{\mathbf{X}}$ is called a *normalized trace evaluation map* and $L_{\mathbf{X}}(1) = \operatorname{tr}(I) = \operatorname{Tr}(I)/d = 1$ holds. A (normalized) trace evaluation map is symmetric, tracial, and positive, which follows from

$$\operatorname{Tr}(X^*) = \operatorname{Tr}(X), \quad \operatorname{Tr}(XY) = \operatorname{Tr}(YX), \quad \operatorname{Tr}(X^*X) \geq 0 \quad \text{for all} \quad X, Y \in \mathcal{H}^d.$$

When $\mathbf{X}$ consists of complex Hermitian matrices, the trace evaluation $L_{\mathbf{X}}$ takes complex values and we may consider its real part $\operatorname{Re}(L_{\mathbf{X}})$, which is also symmetric and tracial, and positive whenever $L_{\mathbf{X}}$ is positive.

The *matrix positivity domain* of a subset $S \subseteq \operatorname{Sym} \mathbb{R}\langle \mathbf{x} \rangle$ of symmetric polynomials is defined as

$$\mathcal{D}(S) = \bigcup_{d \geq 1} \left\{ \mathbf{X} = (X_1, \ldots, X_n) \in (\mathcal{H}^d)^n : g(\mathbf{X}) \succeq 0 \text{ for all } g \in S \right\}. \tag{2}$$

Hence the trace evaluation map $L_{\mathbf{X}}$ for $\mathbf{X} \in \mathcal{D}(S)$ is nonnegative on the *quadratic module*

$$\mathcal{M}(S) = \operatorname{cone}\left\{ p^*gp : p \in \mathbb{R}\langle \mathbf{x} \rangle, \ g \in S \cup \{1\} \right\}.$$

We will now discuss several key results about extending and representing linear forms that are nonnegative on a (truncated) quadratic module. These results are already (mostly) known, but as they will be used extensively in our paper, we provide all proofs in the appendix for the reader's convenience. For the proofs we will work only on the "moment side", which is the natural setting for the applications treated in this paper, thus without using algebraic results about sums of squares of polynomials. In addition we will formulate our results making the link to C^*-algebras explicit whenever relevant, since this will be useful for later results in this paper and also in the work [GdLL17b]. As infinite dimensional analogs of matrix algebras, C^*-algebras arise naturally when considering tracial polynomial optimization problems.

We start with some background on C^*-algebras; see, e.g., [Bla06] for details. For our purposes we define a C^*-*algebra* to be a norm closed $*$-subalgebra over $\mathbb{C}$ of the space $\mathcal{B}(H)$ of bounded operators on a complex Hilbert space H. Such an algebra $\mathcal{A}$ is *unital* if it contains the identity operator (denoted 1). By a fundamental result of Artin-Wedderburn, a finite dimensional C^*-algebra is $*$-isomorphic to a direct sum $\bigoplus_{m=1}^{M} \mathbb{C}^{d_m \times d_m}$ of full complex matrix

algebras [BEK78, Wed64]. In particular, any finite dimensional C^*-algebra is unital. The *positive* elements in a C^*-algebra $\mathcal{A}$ are those of the form a^*a for $a \in \mathcal{A}$, and $\mathcal{A}$ is equipped with a norm $\|\cdot\|$ (induced from the operator norm of $\mathcal{B}(H)$) such that $\|a^*a\| = \|a\|^2$. A *state* τ on a unital C^*-algebra $\mathcal{A}$ is a linear form that is *positive*, i.e., $\tau(a^*a) \geq 0$ for all $a \in \mathcal{A}$, and satisfies $\tau(1) = 1$. Since $\mathcal{A}$ is a complex algebra, every state τ is Hermitian: $\tau(a) = \overline{\tau(a^*)}$ for all $a \in \mathcal{A}$. The state τ is tracial if $\tau(ab) = \tau(ba)$ for all $a, b \in \mathcal{A}$ and *faithful* if $\tau(a^*a) = 0$ implies $a = 0$. A useful fact is that the normalized matrix trace is the unique tracial state on the full matrix algebra $\mathbb{C}^{d \times d}$, see, e.g., [BK12]. Theorems 2.1 and 2.2 use a Gelfand–Naimark–Segal (GNS) construction to represent a positive linear functional as a tracial state on a unital C^*-algebra. Given a unital C^*-algebra $\mathcal{A}$, we need the following adaptation of the matrix positivity domain:

$$\mathcal{D}_{\mathcal{A}}(S) = \left\{ \mathbf{X} \in \mathcal{A}^n : X_i^* = X_i \text{ for all } i \in [n], \, g(\mathbf{X}) \succeq 0 \text{ for all } g \in S \right\},$$

where $g(\mathbf{X}) \succeq 0$ means that $g(\mathbf{X})$ is positive in $\mathcal{A}$.

Analogously to the truncated quadratic module and matrix positivity domain, we will also use truncated ideals and matrix varieties. For $T \subseteq \mathbb{R}\langle \mathbf{x} \rangle$ we define

$$\mathcal{I}_{2t}(T) = \left\{ ph : p \in \mathbb{R}\langle \mathbf{x} \rangle, \, h \in T, \, \deg(ph) \leq 2t \right\},$$
$$\mathcal{V}(T) = \bigcup_{d \geq 1} \left\{ \mathbf{X} \in (\mathcal{H}^d)^n : h(\mathbf{X}) = 0 \text{ for all } h \in T \right\},$$
$$\mathcal{V}_{\mathcal{A}}(T) = \left\{ \mathbf{X} \in \mathcal{A}^n : X_i^* = X_i \text{ for all } i \in [n], \, h(\mathbf{X}) = 0 \text{ for all } h \in T \right\}.$$

When all polynomials $h \in T$ are symmetric one can capture the constraint $L = 0$ on $\mathcal{I}_{2t}(T)$ by requiring $L \geq 0$ on $\mathcal{M}_{2t}(\pm T)$. An important advantage of treating the ideal separately, however, is that we can also use nonsymmetric generators in T.

In the following theorem we assume $\mathcal{M}(S) + \mathcal{I}(T)$ is *Archimedean*, by which we mean that there exists a scalar $R > 0$ such that $R - \sum_{i=1}^n x_i^2 \in \mathcal{M}(S) + \mathcal{I}(T)$. The following theorem is implicit in several works (see, e.g., [NPA12, BKP16]).

Theorem 2.1. *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 $\mathcal{A}$ with tracial state τ and $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S) \cap \mathcal{V}_{\mathcal{A}}(T)$ with

$$L(p) = \tau(p(\mathbf{X})) \quad \text{for all} \quad p \in \mathbb{R}\langle \mathbf{x} \rangle. \tag{3}$$

The next result can be seen as a finite dimensional analogue of the above result, where we do not need $\mathcal{M}(S)$ to be Archimedean, but instead assume the rank of $M(L)$ is finite. In addition to the Gelfand–Naimark–Segal construction, the proof uses Artin–Wedderburn theory. For the unconstrained case the proof of this can be found in [BK12] and in [BKP16, KP16] this result is extended to the constrained case.

Theorem 2.2. *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 $\mathcal{M}(S)$, zero on $\mathcal{I}(T)$, and has $\operatorname{rank}(M(L)) < \infty$;

(2) there is a finite dimensional C^*-algebra $\mathcal{A}$, a tracial state τ, and $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S) \cap \mathcal{V}_{\mathcal{A}}(T)$ satisfying equation (3);

(3) L is a convex combination of normalized trace evaluations at points in $\mathcal{D}(S) \cap \mathcal{V}(T)$.

The following result claims that any *flat* linear functional on a truncated polynomial space can be extended to a linear functional on the full space of polynomials (with the same positivity properties). It is due to Curto and Fialkow [CF96] in the commutative case and extensions to the noncommutative case can be found in [PNA10] (for eigenvalue optimization) and [BK12] (for trace optimization).

Theorem 2.3. *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, δ-flat, nonnegative on $\mathcal{M}_{2t}(S)$, and zero on $\mathcal{I}_{2t}(T)$. Then L extends to a symmetric, tracial, linear form on $\mathbb{R}\langle \mathbf{x} \rangle$ that is nonnegative on $\mathcal{M}(S)$, zero on $\mathcal{I}(T)$, and whose moment matrix has finite rank.*

Combining Theorems 2.2 and 2.3 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 [KP16, Proposition 6.1] (and in [BK12] for the unconstrained case).

Corollary 2.4. *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, δ-flat, nonnegative on $\mathcal{M}_{2t}(S)$, and zero on $\mathcal{I}_{2t}(T)$, then it extends to a conic combination of trace evaluations at elements of $\mathcal{D}(S) \cap \mathcal{V}(T)$.*

2.2 Specialization to the commutative setting

The definitions and results from the previous section can be adapted to the commutative setting.

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_2x_1, x_3x_4)-entry of $M_2(L)$ is equal to its (x_3x_1, x_2x_4)-entry, which does not hold in general in the non-commutative 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 by simultaneously diagonalizing the matrices.

The quadratic module $\mathcal{M}(S)$ and the ideal $\mathcal{I}(T)$ have immediate specializations to the commutative setting. We only need the following two new definitions: the (scalar) positivity domain and scalar variety of sets $S, T \subseteq \mathbb{R}[\mathbf{x}]$ are given by

$$D(S) = \big\{ a \in \mathbb{R}^n : g(a) \geq 0 \text{ for } g \in S \big\} \text{ and } V(T) = \big\{ a \in \mathbb{R}^n : h(a) = 0 \text{ for } h \in T \big\}. \quad (4)$$

We first give the commutative analogue of Theorem 2.1, where we give an additional integral representation in point (3). The equivalence of points (1) and (3) is proved in [Put93] based on Putinar's Positivstellensatz. In the appendix we give a direct proof on the "moment side" using the Gelfand representation.

Theorem 2.5. *Let $S, T \subseteq \mathbb{R}[\mathbf{x}]$ with $\mathcal{M}(S) + \mathcal{I}(T)$ Archimedean. For $L \in \mathbb{R}[\mathbf{x}]^*$, the following are equivalent:*

(1) L is nonnegative on $\mathcal{M}(S)$, zero on $\mathcal{I}(T)$, and $L(1) = 1$;

(2) there exists a unital commutative C^*-algebra $\mathcal{A}$ with a state τ and $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S) \cap \mathcal{V}_{\mathcal{A}}(T)$ such that $L(p) = \tau(p(\mathbf{X}))$ for all $p \in \mathbb{R}[\mathbf{x}]$;

(3) there is a probability measure μ on $D(S) \cap V(T)$ such that

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

The following is the commutative analogue of Theorem 2.2.

Theorem 2.6. *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 $\mathcal{M}(S)$, zero on $\mathcal{I}(T)$, has $\mathrm{rank}(M(L)) < \infty$, and $L(1) = 1$;

(2) there is a finite dimensional commutative C^*-algebra $\mathcal{A}$, a state τ, and $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S) \cap \mathcal{V}_{\mathcal{A}}(T)$ satisfying equation (3);

(3) L is a convex combination of evaluations at points in $D(S) \cap V(T)$.

The next result, due to Curto and Fialkow [CF96], is the commutative analogue of Corollary 2.4.

Theorem 2.7. *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 δ-flat, nonnegative on $\mathcal{M}_{2t}(S)$, and zero on $\mathcal{I}_{2t}(T)$, then L extends to a conic combination of evaluation maps at points in $D(S) \cap V(T)$.*

We will also use 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 2.8. *Let $S, T \subseteq \mathbb{R}[\mathbf{x}]$ such that $\mathcal{M}(S) + \mathcal{I}(T)$ is Archimedean. If $L \in \mathbb{R}[\mathbf{x}]^*$ is nonnegative on $\mathcal{M}(S)$ and zero on $\mathcal{I}(T)$, 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 $D(S) \cap V(T)$.*

2.3 Commutative and tracial polynomial optimization

We briefly recall here the basic polynomial optimization problem in both the commutative and tracial settings. We recall how to design hierarchies of semidefinite programming based bounds and we give their main convergence properties. In the rest of the paper we will see how these results can be adapted to give hierarchies of bounds for matrix factorization ranks. 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 (4):

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

In tracial polynomial optimization, given $f \in \mathrm{Sym}\,\mathbb{R}\langle\mathbf{x}\rangle$, this is modified to minimizing $\mathrm{tr}(f(\mathbf{X}))$ over a feasible region of the form $\mathcal{D}(S)$ as in (2):

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

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

For the commutative case, Lasserre [Las01] and Parrilo [Par00] 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 [PNA10, NPA12] and later to tracial optimization [BCKP13, KP16]. 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 $\mathcal{D}(S)$, and then to express computationally tractable properties satisfied by such maps L.

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

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

We can use this to formulate the following semidefinite programming lower bound on f_*:

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

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

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

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

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

The next theorem from [Las01] gives fundamental convergence properties for the commutative case; see, e.g., [Las09, Lau09] for a detailed exposition.

Theorem 2.9. *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 $\mathcal{M}(S)$ is Archimedean, then $f_t \to f_\infty$ as $t \to \infty$, and the optimal values in f_∞ and f_* are attained and $f_\infty = f_*$.

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

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

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

This problem can be seen as an infinite dimensional analogue of f_*^{tr}: if we restrict to finite dimensional C^*-algebras in the definition of $f_{\mathrm{II}_1}^{\mathrm{tr}}$, then we recover f_*^{tr} (cf. Theorem 2.2). Moreover, under a certain 'flatness' condition, equality $f_*^{\mathrm{tr}} = f_{\mathrm{II}_1}^{\mathrm{tr}}$ holds (cf. Theorem 2.11). Whether $f_{\mathrm{II}_1}^{\mathrm{tr}} = f_*^{\mathrm{tr}}$ is true in general is related to Connes' embedding conjecture (see [KS08, KP16, BKP16]).

Above we defined the parameter $f_{\mathrm{II}_1}^{\mathrm{tr}}$ using C^*-algebras. However, the following lemma shows that we get the same optimal value if we restrict to $\mathcal{A}$ being a von Neumann algebra of type II_1 with separable predual, which is the more common way of defining the parameter $f_{\mathrm{II}_1}^{\mathrm{tr}}$ as is done in [KP16] (and justifies the notation). We omit the proof of this lemma which relies on a GNS construction and algebraic manipulations, standard for algebraists.

Lemma 2.10. *Let $\mathcal{A}$ be a C^*-algebra with tracial state τ and $a_1, \ldots, a_n \in \mathcal{A}$. There exists a von Neumann algebra $\mathcal{F}$ of type II_1 with separable predual, a faithful normal tracial state ϕ, and elements $b_1, \ldots, b_n \in \mathcal{F}$, so that for every $p \in \mathbb{R}\langle \mathbf{x} \rangle$ we have*

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

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

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

$$f_t^{\mathrm{tr}} \leq f_\infty^{\mathrm{tr}} \leq f_{\mathrm{II}_1}^{\mathrm{tr}} \leq f_*^{\mathrm{tr}},$$

where the last inequality follows by considering for $\mathcal{A}$ the full matrix algebra $\mathbb{C}^{d \times d}$ ($d \in \mathbb{N}$). The next theorem from [KP16] summarizes convergence properties for tracial optimization.

Theorem 2.11. *Let $1 \leq \delta \leq t < \infty$ and $S \cup \{f\} \subseteq \mathrm{Sym}\,\mathbb{R}\langle \mathbf{x} \rangle_{2\delta}$ with $\mathcal{D}(S) \neq \emptyset$.*

    (i) If $\mathcal{M}(S)$ is Archimedean, then $f_t^{tr} \to f_\infty^{\mathrm{tr}}$ as $t \to \infty$, and the optimal values in f_∞^{tr} and $f_{\mathrm{II}_1}^{\mathrm{tr}}$ are attained and equal.

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

11

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 2.11(i) and it will be used again later in the paper.

Lemma 2.12. *Let $S \subseteq \operatorname{Sym} \mathbb{R}\langle \mathbf{x} \rangle$ and $R > 0$ such that $R - (x_1^2 + \cdots + x_n^2) \in \mathcal{M}_2(S)$. If $L_t \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ is nonnegative on $\mathcal{M}_{2t}(S)$, then $|L_t(w)| \leq R^{|w|/2} L_t(1)$ for all $w \in \langle \mathbf{x} \rangle_{2t}$. If in addition $\sup_t L_t(1) < \infty$, then (L_t) has a pointwise converging subsequence in $\mathbb{R}\langle \mathbf{x} \rangle^*$.*

3 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 to lower bound its complex completely positive semidefinite rank:

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

where we set

$$S_A^{\mathrm{cpsd}} = \big\{ \sqrt{A_{11}} x_1 - x_1^2, \ldots, \sqrt{A_{nn}} x_n - x_n^2 \big\}. \tag{5}$$

Additionally, define $\xi_*^{\mathrm{cpsd}}(A)$ by adding the constraint $\operatorname{rank}(M(L)) < \infty$ to the program defining $\xi_\infty^{\mathrm{cpsd}}(A)$. This gives a hierarchy of monotone nondecreasing lower bounds on the completely positive semidefinite rank:

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

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

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 $\mathbf{X} \in \mathcal{D}(S)$, and we say that S is *localizing for A* if S is localizing at some factorization $\mathbf{X} \in (\mathcal{H}_+^d)^n$ of A with $d = \operatorname{cpsd-rank}_{\mathbb{C}}(A)$. The set S_A^{cpsd} as defined in (5) is localizing for A. In fact, it is localizing at any factorization $\mathbf{X}$ of A by Hermitian positive semidefinite matrices. Indeed, since $A_{ii} = \operatorname{Tr}(X_i^2) \geq \lambda_{\max}(X_i^2) = \lambda_{\max}(X_i)^2$ we have $\sqrt{A_{ii}} X_i - X_i^2 \succeq 0$ for all $i \in [n]$.

We can use this to show $\xi_*^{\mathrm{cpsd}}(A) \leq \operatorname{cpsd-rank}_{\mathbb{C}}(A)$: Set $d = \operatorname{cpsd-rank}_{\mathbb{C}}(A)$, let $\mathbf{X} \in (\mathcal{H}_+^d)^n$ be a Gram factorization of A and define the linear form $L_{\mathbf{X}} \in \mathbb{R}\langle \mathbf{x} \rangle^*$ by

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

By construction $L_{\mathbf{X}}$ is symmetric and tracial, and $A = (L(x_i x_j))$. Moreover, since the set of polynomials S_A^{cpsd} is localizing for A, the form $L_{\mathbf{X}}$ is nonnegative on $\mathcal{M}(S_A^{\mathrm{cpsd}})$. Finally, we have $\operatorname{rank}(M(L_{\mathbf{X}})) < \infty$, since the algebra generated by $X_1, \ldots, X_n$ is finite dimensional. Hence, $L_{\mathbf{X}}$ is feasible for $\xi_*^{\mathrm{cpsd}}(A)$ with $L_{\mathbf{X}}(1) = d$, which shows $\xi_*^{\mathrm{cpsd}}(A) \leq \operatorname{cpsd-rank}_{\mathbb{C}}(A)$.

As the following inclusions show, the quadratic module $\mathcal{M}(S_A^{\mathrm{cpsd}})$ is Archimedean. Moreover, although there are other possible choices for the localizing polynomials to use in S_A^{cpsd}, the choice made in (5) leads to the largest truncated quadratic module and thus to the best bound. For $c > 0$, we have the inclusions

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

which hold in light of the following identities:

$$c - x = \big((c-x)^2 + c^2 - x^2\big)/(2c), \tag{6}$$

$$c^2 - x^2 = (c-x)^2 + 2(cx - x^2), \tag{7}$$

$$cx - x^2 = \big((c-x)x(c-x) + x(c-x)x\big)/c, \tag{8}$$

$$x = \big((cx - x^2) + x^2\big)/c. \tag{9}$$

In the rest of this section we investigate properties of the above hierarchy as well as 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 $\xi_*^{\text{cpsd}}(A) \le \text{cpsd-rank}_{\mathbb{C}}(A)$ is typically strict, we present an approach to strengthen the bounds in order to go beyond $\xi_*^{\text{cpsd}}(A)$. We also propose some techniques to simplify the computation of the bounds, and we illustrate the behaviour of the bounds on some examples.

3.1 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.

Proposition 3.1. *Let $A \in \mathcal{CS}_+^n$. 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 $\alpha \ge 0$ for which there exists a unital C^*-algebra $\mathcal{A}$ with tracial state τ and $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S_A^{\text{cpsd}})$ 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 $\xi_\infty^{\text{cpsd}}(A)$, which means $\lim_{t \to \infty} \xi_t^{\text{cpsd}}(A)$ exists and is at most $\xi_\infty^{\text{cpsd}}(A)$.

As $\xi_t^{\text{cpsd}}(A) \le \xi_\infty^{\text{cpsd}}(A)$ we may add the constraint $L(1) \le \xi_\infty^{\text{cpsd}}(A)$ to the problem $\xi_t^{\text{cpsd}}(A)$ for every $t \in \mathbb{N}$. By (7) we have $\text{Tr}(A) - \sum_i x_i^2 \in \mathcal{M}_2(S_A^{\text{cpsd}})$. Hence, by Lemma 2.12, the feasible region of $\xi_t^{\text{cpsd}}(A)$ is compact, and thus it has an optimal solution L_t. Again by Lemma 2.12, 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 $\mathcal{M}(S_A^{\text{cpsd}})$. Hence L is feasible for $\xi_\infty^{\text{cpsd}}(A)$. This implies that L is optimal for $\xi_\infty^{\text{cpsd}}(A)$ and that $\lim_{t \to \infty} \xi_t^{\text{cpsd}}(A) = \xi_\infty^{\text{cpsd}}(A)$.

The reformulation of $\xi_\infty^{\text{cpsd}}(A)$ using C^*-algebras follows directly from Theorem 2.1. $\qquad\square$

The following proposition is a direct application of Corollary 2.4. In general we do not know whether the infimum in $\xi_*^{\text{cpsd}}(A)$ is attained, but as the following result shows this infimum is attained if there is a $t \in \mathbb{N}$ for which $\xi_t^{\text{cpsd}}(A)$ admits a flat optimal solution.

Proposition 3.2. *Let $A \in \mathcal{CS}_+^n$. The parameter $\xi_*^{\text{cpsd}}(A)$ is equal to the infimum of $L(1)$ over all conic combinations L of trace evaluations at elements of $\mathcal{D}_{\mathcal{A}}(S_A^{\text{cpsd}})$ such that $A = (L(x_i x_j))$. Equivalently, $\xi_*^{\text{cpsd}}(A)$ is the infimum over all $\alpha \ge 0$ for which there exist a finite dimensional C^*-algebra $\mathcal{A}$ with tracial state τ and*

$$(X_1, \ldots, X_n) \in \mathcal{D}_{\mathcal{A}}(S_A^{\text{cpsd}}) \quad \text{such that} \quad A = \alpha \cdot (\tau(X_i X_j)).$$

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

We have the following more explicit formulation for $\xi_*^{\text{cpsd}}(A)$, which also explains why the inequality $\xi_*^{\text{cpsd}}(A) \le \text{cpsd-rank}_{\mathbb{C}}(A)$ is typically strict. Here $\| \cdot \|$ denotes the operator norm, so that $\|X\| = \lambda_{\max}(X)$ for $X \succeq 0$.

Proposition 3.3. *For $A \in \mathcal{CS}_+^n$ we have*

$$\xi_*^{\mathrm{cpsd}}(A) = \inf\Big\{ \sum_{m=1}^{M} d_m \cdot \max_{i \in [n]} \frac{\|X_i^m\|^2}{A_{ii}} : M \in \mathbb{N}, d_1, \ldots, d_M \in \mathbb{N}, \tag{10}$$
$$X_i^m \in \mathcal{H}_+^{d_m} \text{ for } i \in [n], m \in [M],$$
$$A = \mathrm{Gram}\big(\oplus_{m=1}^{M} X_1^m, \ldots, \oplus_{m=1}^{M} X_n^m \big)\Big\}.$$

Note that using matrices from $\mathcal{S}_+^{d_m}$ instead of $\mathcal{H}_+^{d_m}$ 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 $\mathcal{D}(S_A^{\mathrm{cpsd}})$ as given in Proposition 3.2. We first show the inequality $\beta \leq \xi_*^{\mathrm{cpsd}}(A)$, where β is the optimal value of the program in (10).

For this, assume $L \in \mathbb{R}\langle \mathbf{x} \rangle^*$ is a conic combination of trace evaluations at elements of $\mathcal{D}(S_A^{\mathrm{cpsd}})$ such that $A = (L(x_i x_j))$. We will construct a feasible solution for (10) with objective value $L(1)$. The linear functional L can be written as

$$L = \sum_{m=1}^{M} \lambda_m L_{\mathbf{Y}^m}, \quad \text{where} \quad \lambda_m > 0 \quad \text{and} \quad \mathbf{Y}^m \in \mathcal{D}(S_A^{\mathrm{cpsd}}) \quad \text{for all} \quad m \in [M].$$

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 $\mathbf{Y}^m \in \mathcal{D}(S_A^{\mathrm{cpsd}})$, we have $Y_i^m \succeq 0$ and $A_{ii}I - (Y_i^m)^2 \succeq 0$ by identities (7) and (9). This implies $\|Y_i^m\|^2 \leq A_{ii}$ for all $i \in [n]$ and $m \in [M]$. Let $\mathbf{X}^m = \sqrt{\lambda_m} \mathbf{Y}^m$. Then, $L(x_i x_j) = \sum_m \mathrm{Tr}(X_i^m X_j^m)$, 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 (10) with value

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

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

For the other direction we assume

$$A = \mathrm{Gram}\big(\oplus_{m=1}^{M} X_1^m, \ldots, \oplus_{m=1}^{M} X_n^m \big), \quad X_1^m, \ldots, X_n^m \in \mathcal{S}_+^{d_m} \ (m \in [M]).$$

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 $\mathcal{D}(S_A^{\mathrm{cpsd}})$. For this we observe that $\lambda_m A_{ii} \geq \|X_i^m\|^2$. Therefore $\lambda_m A_{ii}I \succeq (X_i^m)^2$, so $A_{ii}I \succeq (Y_i^m)^2$, which implies $\sqrt{A_{ii}}Y_i^m - (Y_i^m)^2 \succeq 0$. This shows $\xi_*^{\mathrm{cpsd}}(A) \leq L(1) = \sum_m \lambda_m d_m$, and thus $\xi_*^{\mathrm{cpsd}}(A) \leq \beta$. $\qquad\square$

We can say a bit more when A lies on an extreme ray of the cone $\mathcal{CS}_+^n$: In the formulation from Proposition 3.3 it suffices to restrict the minimization over all factorizations of A involving only one block. We know very little about the extreme rays of $\mathcal{CS}_+^n$, also in view of the recent result [Slo17] that the cone $\mathcal{CS}_+^n$ is not closed for large n.

Proposition 3.4. *If A lies on an extreme ray of the cone $\mathcal{CS}_+^n$, then*

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

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 (10) and some block X_i^m has rank 1, then $\xi_*^{\mathrm{cpsd}}(A) = \mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A)$.

Proof. Let β denote the infimum in Proposition 3.4. The inequality $\xi_*^{\mathrm{cpsd}}(A) \leq \beta$ follows from the reformulation of $\xi_*^{\mathrm{cpsd}}(A)$ in Proposition 3.3. 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 (10), 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 $A_m = \mathrm{Gram}(X_1^m, \cdots, X_n^m)$, so that $A = \sum_m A_m$. As A lies on an extreme ray of $\mathcal{CS}_+^n$, we must have $A_m = \alpha_m A$ for some $\alpha_m > 0$ with $\sum_m \alpha_m = 1$. Hence, since

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

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 $\beta \leq \xi_*^{\mathrm{cpsd}}(A)$, 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 (10) and that there is a block X_i^m of rank 1. By Proposition 3.3 we have $\sum_m d_m \lambda_m = \xi_*^{\mathrm{cpsd}}(A)$. From the argument just made above it follows that

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

As $\sum_m \alpha_m = 1$ this implies equality $d_m \lambda_m / \alpha_m = \min_m d_m \lambda_m / \alpha_m$ for all m, that is, all terms $d_m \lambda_m / \alpha_m$ take the same value, equal to $\xi_*^{\mathrm{cpsd}}(A)$. 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 $\xi_*^{\mathrm{cpsd}}(A) = d_m$. On the other hand, $\mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A) \leq d_m$ since $(X_i^m / \sqrt{\alpha_m})_i$ forms a Gram decomposition of A, and thus equality $\xi_*^{\mathrm{cpsd}}(A) = d_m = \mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A)$ holds. $\qquad\square$

3.2 Additional localizing constraints to improve on $\xi_*^{\mathrm{cpsd}}(A)$

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 3.5. *Let $A \in \mathcal{CS}_+^n$. For $v \in \mathbb{R}^n$, the polynomial $g_v = v^{\mathsf{T}} A v - \left(\sum_{i=1}^n v_i x_i \right)^2$ is localizing for A (at any Gram factorization by Hermitian positive semidefinite matrices).*

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

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

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

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

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

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

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

By scaling invariance, we can add all of the above constraints (for all $v \in \mathbb{R}^n$) by setting V to be the unit sphere S^{n-1}. Since S^{n-1} is a compact metric space, there exists a sequence $V_1 \subseteq V_2 \subseteq \ldots \subseteq S^{n-1}$ of finite subsets such that $\bigcup_{k \geq 1} V_k$ is dense in S^{n-1}.

15

Proposition 3.6. *Consider a matrix $A \in \mathcal{CS}_+^n$. For $t \in \{\infty, *\}$, we have*

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

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

$$\|u - v\|_1 \leq \frac{\varepsilon \lambda_{\min}(A)}{4\sqrt{n} \max_i A_{ii}} \quad \text{and} \quad \|u - v\|_2 \leq \frac{\varepsilon \lambda_{\min}(A)}{4 \mathrm{Tr}(A^2)^{1/2}}. \tag{11}$$

Propositions 3.1 and 3.2 have analogues for the programs $\xi_{t,V}^{\mathrm{cpsd}}(A)$. These show that $\xi_{t,V_k}^{\mathrm{cpsd}}(A)$, for $t = \infty$ ($t = *$), is the infimum over all $\alpha \geq 0$ for which there exist a (finite dimensional) unital C^*-algebra $\mathcal{A}$ with tracial state τ and $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S_{A,V_k}^{\mathrm{cpsd}})$ such that $A = \alpha \cdot (\tau(X_i X_j))$.

Below we show that
$$\mathbf{X}' = \sqrt{1 - \varepsilon} \mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S_{A,S^{n-1}}^{\mathrm{cpsd}}).$$

This implies that $L \in \mathbb{R}\langle \mathbf{x} \rangle^*$ defined by $L(p) = \alpha/(1 - \varepsilon)\tau(p(\mathbf{X}'))$ is feasible for $\xi_{t,S^{n-1}}^{\mathrm{cpsd}}(A)$ with objective value $L(1) = \alpha/(1 - \varepsilon)$. Since $\varepsilon > 0$ was arbitrary, this completes the proof.

Consider the map
$$f_{\mathbf{X}} \colon S^{n-1} \to \mathbb{R}, \; v \mapsto \Big\| \sum_{i=1}^n v_i X_i \Big\|^2,$$

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

$$v^\mathsf{T} A v - f_{\mathbf{X}'}(v) = v^\mathsf{T} A v \Big(1 - (1 - \varepsilon) \frac{f_{\mathbf{X}}(v)}{v^\mathsf{T} A v} \Big) \geq v^\mathsf{T} A v \big(1 - (1 - \varepsilon) \big) = \varepsilon v^\mathsf{T} A v \geq \varepsilon \lambda_{\min}(A).$$

Let $u \in S^{n-1}$ and let $v \in V_k$ be such that (11) 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 $\mathcal{A}$, we have that $\sqrt{A_{ii}} - X_i$ is positive in $\mathcal{A}$ by (6) and (7), which implies $\|X_i\| \leq \sqrt{A_{ii}}$. By the reverse triangle inequality we then have

$$|f_{\mathbf{X}'}(u) - f_{\mathbf{X}'}(v)| = \Big| \Big\| \sum_{i=1}^n u_i X_i' \Big\| - \Big\| \sum_{i=1}^n v_i X_i' \Big\| \Big| \Big(\Big\| \sum_{i=1}^n u_i X_i' \Big\| + \Big\| \sum_{i=1}^n v_i X_i' \Big\| \Big)$$
$$\leq \Big\| \sum_{i=1}^n (v_i - u_i) X_i' \Big\| 2\sqrt{n} \max_i \sqrt{A_{ii}}$$
$$\leq \Big(\sum_{i=1}^n |v_i - u_i| \|X_i'\| \Big) 2\sqrt{n} \max_i \sqrt{A_{ii}}$$
$$\leq \|u - v\|_1 2\sqrt{n} \max_i A_{ii} \leq \frac{\varepsilon \lambda_{\min}(A)}{4\sqrt{n} \max_i A_{ii}} 2\sqrt{n} \max_i A_{ii} = \frac{\varepsilon \lambda_{\min}(A)}{2}.$$

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

16

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

Example 3.7. Consider the matrix

$$A = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} = \mathrm{Gram}\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \right), \tag{12}$$

with cpsd-rank$_{\mathbb{C}}(A) = 2$. We can also write $A = \mathrm{Gram}(Y_1, Y_2)$, 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 $I - X_i^2 \succeq 0$. Hence the linear form $L = L_{\mathbf{X}}/2$ is feasible for $\xi_*^{\mathrm{cpsd}}(A)$, which shows that $\xi_*^{\mathrm{cpsd}}(A) \leq L(1) = 3/2$. In fact, this form L gives an optimal flat solution to $\xi_2^{\mathrm{cpsd}}(A)$, as we can check using a semidefinite programming solver, and hence $\xi_*^{\mathrm{cpsd}}(A) = 3/2$. In passing, we observe that $\xi_1^{\mathrm{cpsd}}(A) = 4/3$, which coincides with the analytic lower bound (15) (see also Lemma 3.13 below).

For $e = (1,1) \in \mathbb{R}^2$ and $V = \{e\}$, this form L is not feasible for $\xi_{*,V}^{\mathrm{cpsd}}(A)$, because for $p = 1 - 3x_1 - 3x_2$ 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 each $t \geq 2$ it cuts off part of the feasibility region of $\xi_t^{\mathrm{cpsd}}(A)$. Indeed, using a semidefinite programming solver we find an optimal flat solution of $\xi_{3,V}^{\mathrm{cpsd}}(A)$ with objective value $(5 - \sqrt{3})/2 \approx 1.633$, which shows

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

Example 3.8. Consider the symmetric circulant matrices

$$M(\alpha) = \begin{pmatrix} 1 & \alpha & 0 & 0 & \alpha \\ \alpha & 1 & \alpha & 0 & 0 \\ 0 & \alpha & 1 & \alpha & 0 \\ 0 & 0 & \alpha & 1 & \alpha \\ \alpha & 0 & 0 & \alpha & 1 \end{pmatrix}.$$

For $0 \leq \alpha \leq 1/2$ we have $M(\alpha) \in \mathcal{CS}_+^5$ with 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

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

form a factorization of $M(\alpha)$. As $M(\alpha)$ is supported by a cycle, we have $M(\alpha) \in \mathcal{CS}_+^5$ if and only if $M(\alpha) \in \mathcal{CP}^5$ and thus $M(\alpha) \in \mathcal{CS}_+^5$ if and only if $0 \leq \alpha \leq 1/2$ [LP15].

By using the formulation in Proposition 3.3, we can use the above factorization to derive $\xi_*^{\mathrm{cpsd}}(M(1/2)) \leq 5/2$. However, using a semidefinite programming solver we see that

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

where V is the set containing the vector $(1, -1, 1, -1, 1)$ and its cyclic shifts. Hence the bound $\xi_{2,V}^{\mathrm{cpsd}}(M(1/2))$ is tight: It certifies cpsd-rank$_{\mathbb{C}}(M(1/2)) = 5$, while the other known bounds, the rank bound $\sqrt{\mathrm{rank}(A)}$ and the analytic bound (15), 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 $\xi_1^{\mathrm{cpsd}}(M(0)) = 5$ (by Lemma 3.13), the above result $\xi_{2,V}^{\mathrm{cpsd}}(M(1/2)) = 5$, and the lower semicontinuity of $\alpha \mapsto \xi_{2,V}^{\mathrm{cpsd}}(M(\alpha))$ shown in Lemma 3.14. 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. $\qquad \triangle$

3.3 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 sometimes shrink the feasibility region of $\xi_{t,V}^{\mathrm{cpsd}}(A)$ for $t \in \mathbb{N}$, but they are redundant for the parameters $\xi_{\infty,V}^{\mathrm{cpsd}}(A)$ and $\xi_{*,V}^{\mathrm{cpsd}}(A)$. The latter is shown using the reformulation of $\xi_{t,V}^{\mathrm{cpsd}}(A)$ ($t \in \{\infty, *\}$) 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 [BFS16]. Note that as for localizing constraints, these bilinear constraints can be modeled as semidefinite constraints.

Lemma 3.9. *Let $A \in \mathcal{CS}_+^n$, $t \in \mathbb{N} \cup \{\infty, *\}$, and let g, g' be localizing for A. If we add the constraints*

$$L(p^* g p g') \geq 0 \quad for \quad p \in \mathbb{R}\langle \mathbf{x} \rangle \quad with \quad \deg(p^* g p g') \leq 2t \tag{13}$$

to $\xi_{t,V}^{\mathrm{cpsd}}(A)$, then it still lower bounds $\mathrm{cpsd\text{-}rank}_{\mathbb{C}}(A)$. Moreover, if $g, g' \in \mathcal{M}(S_{A,V}^{\mathrm{cpsd}})$ then the constraints (13) *are redundant for $\xi_{\infty,V}^{\mathrm{cpsd}}(A)$ and $\xi_{*,V}^{\mathrm{cpsd}}(A)$.*

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

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

Now suppose that L is feasible for $\xi_{t,V}^{\mathrm{cpsd}}(A)$ with $t \in \{\infty, *\}$. By Theorem 2.1 there exist a unital C^*-algebra $\mathcal{A}$ with tracial state τ and $\mathbf{X} \in \mathcal{D}(S_{A,V}^{\mathrm{cpsd}})$ such that $L(p) = L(1)\tau(p(\mathbf{X}))$ for all $p \in \mathbb{R}\langle \mathbf{x} \rangle$. Since $g, g' \in \mathcal{M}(S_{A,V}^{\mathrm{cpsd}})$ we know that $g(\mathbf{X}), g'(\mathbf{X})$ are positive elements in $\mathcal{A}$, that is, $g(\mathbf{X}) = a^* a$ and $g'(\mathbf{X}) = b^* b$ for some $a, b \in \mathcal{A}$. Then we have

$$
\begin{aligned}
L(p^* g p g) &= L(1)\, \tau(p^*(\mathbf{X})\, g(\mathbf{X})\, p(\mathbf{X})\, g'(\mathbf{X})) \\
&= L(1)\, \tau(p^*(\mathbf{X})\, a^* a\, p(\mathbf{X})\, b^* b) \\
&= L(1)\, \tau((a\, p(\mathbf{X})\, b^*)^* a\, p(\mathbf{X})\, b^*) \geq 0
\end{aligned}
$$

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

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

Lemma 3.10. *Let $A \in \mathcal{CS}_+^n$ and $t \in \mathbb{N} \cup \{\infty, *\}$. If we add the constraints*

$$L = 0 \quad on \quad \mathcal{I}_{2t}\Big(\Big\{\sum_{i=1}^n v_i x_i : v \in \ker A\Big\} \cup \big\{x_i x_j : A_{ij} = 0\big\}\Big) \tag{14}$$

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

Proof. Let $\mathbf{X} \in (\mathcal{H}_+^d)^n$ be a Gram decomposition of A, and let L be the real part of the trace evaluation $L_{\mathbf{X}}$. If $Av = 0$, then $0 = v^{\mathsf{T}} A v = \mathrm{Tr}((\sum_{i=1}^n v_i X_i)^2)$, which implies $\sum_{i=1}^n v_i X_i = 0$ and thus $L((\sum_{I=1}^n v_i x_i)p) = \mathrm{Re}(\mathrm{Tr}((\sum_{i=1}^n v_i X_i)p(\mathbf{X}))) = 0$.

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. It follows that $L(x_i x_j p) = \mathrm{Re}(\mathrm{Tr}(X_i X_j p(\mathbf{X}))) = 0$.

As in the proof of the previous lemma, if L is feasible for $\xi_{t,V}^{\mathrm{cpsd}}(A)$ with $t \in \{\infty, *\}$ then, by Theorem 2.1, there exist a unital C^*-algebra $\mathcal{A}$ with tracial state τ and $\mathbf{X} \in \mathcal{D}(S_{A,V}^{\mathrm{cpsd}})$ such that $L(p) = L(1)\tau(p(\mathbf{X}))$ for all $p \in \mathbb{R}\langle \mathbf{x} \rangle$. Moreover, by Lemma 2.10 we may assume τ to be faithful. If $v \in \ker A$ then $0 = v^{\mathsf{T}} A v = L((\sum_i v_i x_i)^2) = L(1)\tau((\sum_i v_i X_i)^2)$, and hence, since

τ is faithful, $\sum_i v_i X_i = 0$ in $\mathcal{A}$. 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, since X_i, X_j are positive in $\mathcal{A}$ and τ is faithful, thus $X_i X_j = 0$. From which it follows that $L(p x_i x_j) = 0$ for all $p \in \mathbb{R}\langle \mathbf{x}\rangle$. $\qquad\square$

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 (14), are in fact redundant, because, 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 (14), it is more efficient to index the moment matrices with a basis for $\mathbb{R}\langle \mathbf{x}\rangle_t$ modulo the ideal $\mathcal{I}_t\big(\{\sum_i v_i x_i : v \in \ker(A)\} \cup \{x_i x_j : A_{ij} = 0\}\big)$.

3.4 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 3.11. *Suppose $A \in \mathcal{CS}_+^n$ and $t \in \mathbb{N} \cup \{\infty, *\}$.*

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

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

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

We also have the following direct sum property, where the equality follows using the C^*-algebra reformulations as given in Proposition 3.1 and Proposition 3.2.

Lemma 3.12. *If $A \in \mathcal{CS}_+^n$ and $B \in \mathcal{CS}_+^m$, then $\xi_t^{\mathrm{cpsd}}(A \oplus B) \leq \xi_t^{\mathrm{cpsd}}(A) + \xi_t^{\mathrm{cpsd}}(B)$, where equality holds for $t \in \{\infty, *\}$.*

Proof. To prove the inequality we take L_A and L_B feasible for $\xi_t^{\mathrm{cpsd}}(A)$ and $\xi_t^{\mathrm{cpsd}}(B)$, and construct a feasible L for $\xi_t^{\mathrm{cpsd}}(A \oplus B)$ 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 3.1 (Proposition 3.2), $\xi_t^{\mathrm{cpsd}}(A \oplus B)$ is equal to the smallest $\alpha \geq 0$ for which there exists a (finite dimensional) unital C^*-algebra $\mathcal{A}$ with tracial state τ and $(\mathbf{X}, \mathbf{Y}) \in \mathcal{D}_\mathcal{A}(S_{A \oplus B}^{\mathrm{cpsd}})$ 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 $\mathbf{X} \in \mathcal{D}_\mathcal{A}(S_A^{\mathrm{cpsd}})$ and $\mathbf{Y} \in \mathcal{D}_\mathcal{A}(S_B^{\mathrm{cpsd}})$. Let P_A be the projection onto the space $\sum_i \mathrm{Im}(X_i)$ and define $L_A \in \mathbb{R}\langle \mathbf{x}\rangle^*$ by $L_A(p) = \alpha \cdot \tau(p(\mathbf{X}) P_A)$. It follows that L_A is is nonnegative on $\mathcal{M}(S_A^{\mathrm{cpsd}})$, 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 $\xi_\infty^{\mathrm{cpsd}}(A)$ with $L_A(1) = \alpha \tau(P_A)$. In the same way we consider the projection P_B onto the space $\sum_j \mathrm{Im}(Y_j)$ and define a feasible solution L_B for $\xi_t^{\mathrm{cpsd}}(B)$ with $L_B(1) = \alpha \tau(P_B)$. By Lemma 2.10 we may assume τ 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 $\sum_i \mathrm{Im}(X_i) \perp \sum_j \mathrm{Im}(Y_j)$. 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 $\xi_t^{\mathrm{cpsd}}(A) + \xi_t^{\mathrm{cpsd}}(B) \leq L_A(1) + L_B(1) \leq \alpha$, and thus $\xi_t^{\mathrm{cpsd}}(A) + \xi_t^{\mathrm{cpsd}}(B) \leq \xi_t^{\mathrm{cpsd}}(A \oplus B)$, completing the proof. $\qquad\square$

Note that the completely positive semidefinite rank of a matrix satisfies the same properties as those mentioned in the above two lemmas, where the inequality in Lemma 3.12 is always an equality: $\mathrm{cpsd\text{-}rank}_\mathbb{C}(A \oplus B) = \mathrm{cpsd\text{-}rank}_\mathbb{C}(A) + \mathrm{cpsd\text{-}rank}_\mathbb{C}(B)$ [PSVW16, GdLL17a].

The following lemma shows that the first level of our hierarchy is at least as good as the analytic lower bound (15) on the cpsd-rank derived in [PSVW16, Theorem 10].

Lemma 3.13. *For any non-zero matrix $A \in \mathcal{CS}_+^n$ we have*

$$\xi_1^{\mathrm{cpsd}}(A) \geq \frac{\left(\sum_{i=1}^n \sqrt{A_{ii}}\right)^2}{\sum_{i,j=1}^n A_{ij}}. \tag{15}$$

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. $\square$

As an application of Lemma 3.13, the first bound ξ_1^{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 with Lemma 3.11, 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 5×5 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 [FGP$^+$15] 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 [PSVW16] also evaluates to 2, hence Lemma 3.13 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 $A \in \mathcal{CS}_+^n$ 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 3.14. *For every $t \in \mathbb{N} \cup \{\infty\}$ and $V \subseteq \mathbb{R}^n$, the function*

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

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 $\{A \in \mathcal{S}^n : \xi_{t,V}^{\mathrm{cpsd}}(A) \leq r\}$ are closed. For this we consider a sequence $(A_k)_{k \in \mathbb{N}}$ of symmetric matrices converging to $A \in \mathcal{S}^n$ 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 2.12 that there is a pointwise converging subsequence of (L_k), still denoted (L_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

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

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

Hence we have

$$
\begin{aligned}
L(p^*(\sqrt{A_{ii}}x_i - x_i^2)p) &= \sqrt{A_{ii}}\Big(L(p^*x_i p) - L_k(p^*x_i p) + L_k(p^*x_i p)\Big) \\
&\quad - \Big(L(p^*x_i^2 p) - L_k(p^*x_i^2 p) + L_k(p^*x_i^2 p)\Big) \\
&\geq -2\varepsilon + \sqrt{A_{ii}}\, L_k(p^*x_i p) - L_k(p^*x_i^2 p) \\
&\geq -3\varepsilon + \sqrt{(A_k)_{ii}}\, L_k(p^*x_i p) - L_k(p^*x_i^2 p) \\
&= -3\varepsilon + L_k(p^*(\sqrt{(A_k)_{ii}}\, x_i - x_i^2)p) \geq -3\varepsilon,
\end{aligned}
$$

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

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

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

Lemma 3.15. *For every $t \in \mathbb{N} \cup \{\infty\}$, the function*

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

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

Proof. Let $A, B \in \mathcal{CS}_+^n \cap \mathcal{E}_n$ 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 $L = \lambda L_A + (1 - \lambda)L_B$ is feasible for $\xi_t^{\mathrm{cpsd}}(\lambda A + (1 - \lambda)B)$, which implies

$$\xi_t^{\mathrm{cpsd}}(\lambda A + (1 - \lambda)B) \leq L(1) = \lambda L_A(1) + (1 - \lambda)L_B(1) = \lambda\xi_t^{\mathrm{cpsd}}(A) + (1 - \lambda)\xi_t^{\mathrm{cpsd}}(B). \quad \square$$

Example 3.16. In this example we show that for $t \geq 1$, the function

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

is not continuous. For this we consider the matrices

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

with $\text{cpsd-rank}_{\mathbb{C}}(A_k) = 2$ 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 $\mathcal{CS}_+^n$ 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 (15) of [PSVW16]. $\qquad \triangle$

Example 3.17. In this example we determine $\xi_t^{\mathrm{cpsd}}(A)$ for all $t \geq 1$ and $A \in \mathcal{CS}_+^2$. In view of Lemma 3.11(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 $\xi_1^{\mathrm{cpsd}}(A(\alpha))$ is equal to the analytic bound $2/(\alpha + 1)$ from (15), where the equality follows from the fact that L given by $L(x_ix_j) = A(\alpha)_{ij}$, $L(x_1) = L(x_2) = 1$ and $L(1) = 2/(\alpha + 1)$ is feasible for $\xi_1^{\mathrm{cpsd}}(A(\alpha))$.

For $t \geq 2$ we show $\xi_t^{\mathrm{cpsd}}(A(\alpha)) = 2 - \alpha$. By the above this is true for $\alpha = 0$ and $\alpha = 1$, and in Example 3.7 we show $\xi_t^{\mathrm{cpsd}}(A(1/2)) = 3/2$ for $t \geq 2$. The claim then follows since the function $\alpha \mapsto \xi_t^{\mathrm{cpsd}}(A(\alpha))$ is convex by Lemma 3.15. $\qquad \triangle$

4 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 [FP16]. Their approach relies on the atomicity of the completely positive rank, that is, the fact that $\text{cp-rank}(A) = r$ 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 $\text{cp-rank}(A) = r$, 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

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

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

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

as an efficiently computable relaxation of $\tau_{\mathrm{cp}}(A)$, and they show rank$(A) \leq \tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$, so that

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

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 Section 3 to the commutative setting. Indeed, we may view a factorization $A = (a_i^{\mathsf{T}} a_j)$ 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 pin point the link to the parameters $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$ and $\tau_{\mathrm{cp}}(A)$. The direct analogue of $\xi_t^{\mathrm{cpsd}}(A)$ in the commutative setting leads to a hierarchy that does not converge to $\tau_{\mathrm{cp}}(A)$, but we provide two approaches to strengthen it that do converge to $\tau_{\mathrm{cp}}(A)$. The first approach is based on a generalization of the tensor constraints in $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$. Here 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 $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$ while being defined by a smaller semidefinite program. The second approach relies on adding localizing constraints for vectors in the unit sphere as in Section 3.2.

The following hierarchy is a commutative analogue of the hierarchy from Section 3, where we may now add the localizing polynomials $A_{ij} - x_i x_j$ for all $1 \leq i < j \leq n$, 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

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

where we set

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

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

$$L(gu) \geq 0 \quad \text{for} \quad g \in \{1\} \cup S_A^{\mathrm{cp}} \quad \text{and} \quad u \in [\mathbf{x}]_{2t-\deg(g)} \tag{16}$$

and the tensor constraints

$$(L((ww')^c))_{w,w' \in \langle \mathbf{x} \rangle_{=l}} \preceq A^{\otimes l} \quad \text{for all integers} \quad 2 \leq l \leq t, \tag{17}$$

which generalize the case $l = 2$ used in the relaxation $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$. 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 (17) involve matrices indexed by the *noncommutative* words of length exactly l. In Section 4.4 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 Section 3.2. 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 Section 4.5.2 for an example).

The bounds ξ_t^{cp} and $\xi_{t,\dagger}^{\mathrm{cp}}$ 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 4.1 and 4.2 we show

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

and in Proposition 4.5 we show the equality $\xi_{*,\dagger}^{\mathrm{cp}}(A) = \tau_{\mathrm{cp}}(A)$.

4.1 Comparison to $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$

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

Proposition 4.1. *For $A \in \mathcal{CP}^n$ and $t \in \mathbb{N} \cup \{\infty, *\}$ we have $\xi_{t,\dagger}^{\mathrm{cp}}(A) \leq \tau_{\mathrm{cp}}(A)$.*

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.

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

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

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 [FP16] to show $X_l \preceq A^{\otimes l}$ for all $l \geq 2$. 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

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

Now we show that the new parameter $\xi_{2,\dagger}^{\mathrm{cp}}(A)$ is at least as good as $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$. Later in Section 4.5.1 we will give an example where the inequality is strict.

Proposition 4.2. *For $A \in \mathcal{CP}^n$ we have $\tau_{\mathrm{cp}}^{\mathrm{sos}}(A) \leq \xi_{2,\dagger}^{\mathrm{cp}}(A)$.*

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

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

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. Indeed, 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 (16)), and for $i = j$ this follows from

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

which holds because of (7), 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 $X \preceq A \otimes A$. $\qquad\square$

4.2 Convergence of the basic hierarchy

We first summarize convergence properties of the hierarchy $\xi_t^{\mathrm{cp}}(A)$. Note that unlike in Section 3 where we can only claim the inequality $\xi_\infty^{\mathrm{cpsd}}(A) \le \xi_*^{\mathrm{cpsd}}(A)$, here we can show $\xi_\infty^{\mathrm{cp}}(A) = \xi_*^{\mathrm{cp}}(A)$.

Proposition 4.3. *Let $A \in \mathcal{CP}^n$. For every $t \in \mathbb{N} \cup \{\infty, *\}$ the optimum in $\xi_t^{\mathrm{cp}}(A)$ is attained, and $\xi_t^{\mathrm{cp}}(A) \to \xi_\infty^{\mathrm{cp}}(A) = \xi_*^{\mathrm{cp}}(A)$. If $\xi_t^{\mathrm{cp}}(A)$ admits a flat optimal solution, then $\xi_t^{\mathrm{cp}}(A) = \xi_\infty^{\mathrm{cp}}(A)$. Moreover, $\xi_\infty^{\mathrm{cp}}(A) = \xi_*^{\mathrm{cp}}(A)$ is the minimum value of $L(1)$ taken over all conic combinations L of evaluations at elements of $D(S_A^{\mathrm{cp}})$ satisfying $A = (L(x_i x_j))$.*

Proof. We may assume $A \ne 0$. Since $\sqrt{A_{ii}} x_i - x_i^2 \in S_A^{\mathrm{cp}}$ for all i, using (7) we obtain that $\mathrm{Tr}(A) - \sum_i x_i^2 \in \mathcal{M}_2(S_A^{\mathrm{cp}})$. By adapting the proof of Proposition 3.1 to the commutative setting, we see that the optimum in $\xi_t^{\mathrm{cp}}(A)$ is attained for $t \in \mathbb{N} \cup \{\infty\}$, and $\xi_t^{\mathrm{cp}}(A) \to \xi_\infty^{\mathrm{cp}}(A)$.

We now show the inequality $\xi_*^{\mathrm{cp}}(A) \le \xi_\infty^{\mathrm{cp}}(A)$, which implies that equality holds. For this, let L be optimal for $\xi_\infty^{\mathrm{cp}}(A)$. By Theorem 2.8, the restriction of L to $\mathbb{R}[\mathbf{x}]_2$ extends to a conic combination of evaluations at points in $D(S_A^{\mathrm{cp}})$. It follows that this extension is feasible for $\xi_*^{\mathrm{cp}}(A)$ with the same objective value. This shows that $\xi_*^{\mathrm{cp}}(A) \le \xi_\infty^{\mathrm{cp}}(A)$, that the optimum in $\xi_*^{\mathrm{cp}}(A)$ is attained, and that $\xi_*^{\mathrm{cp}}(A)$ is the minimum of $L(1)$ over all conic combinations L of evaluations at elements of $D(S_A^{\mathrm{cp}})$ such that $A = (L(x_i x_j))$. Finally, by Theorem 2.7 we have $\xi_t^{\mathrm{cp}}(A) = \xi_\infty^{\mathrm{cp}}(A)$ if $\xi_t^{\mathrm{cp}}(A)$ admits a flat optimal solution. $\qquad\square$

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

Proposition 4.4. *We have*

$$\xi_*^{\mathrm{cp}}(A) = \min\Big\{ \alpha : \alpha \ge 0,\ A \in \alpha \cdot \mathrm{conv}\{R \in \mathcal{S}^n : 0 \le R \le A,\ \mathrm{rank}(R) \le 1\} \Big\}.$$

Proof. This follows directly from the reformulation of $\xi_*^{\mathrm{cp}}(A)$ in Proposition 4.3 in terms of conic evaluations at points in $D(S_A^{\mathrm{cp}})$ after observing that, for $v \in \mathbb{R}^n$, we have $v \in D(S_A^{\mathrm{cp}})$ if and only if the matrix $R = vv^{\mathsf{T}}$ satisfies $0 \le R \le A$. $\qquad\square$

4.3 Additional constraints and convergence to $\tau_{\mathrm{cp}}(A)$

The reformulation of the parameter $\xi_*^{\mathrm{cp}}(A)$ in Proposition 4.4 differs from $\tau_{\mathrm{cp}}(A)$ in that the constraint $R \preceq A$ is missing. In order to have a hierarchy converging to $\tau_{\mathrm{cp}}(A)$ we need to add constraints to enforce that L can be decomposed as a conic combination of evaluation maps at nonnegative vectors v satisfying $vv^{\mathsf{T}} \preceq A$. Here we present two ways to achieve this goal. First we show that the tensor constraints (17) suffice in the sense that $\xi_{*,\dagger}^{\mathrm{cp}}(A) = \tau_{\mathrm{cp}}(A)$ (note that the constraints (16) are not needed for this result). However, because of the special form of the tensor constraints we do not know whether $\xi_{t,\dagger}^{\mathrm{cp}}(A)$ admitting a flat optimal solution implies $\xi_{t,\dagger}^{\mathrm{cp}}(A) = \xi_{*,\dagger}^{\mathrm{cp}}(A)$, and we do not know whether $\xi_{\infty,\dagger}^{\mathrm{cp}}(A) = \xi_{*,\dagger}^{\mathrm{cp}}(A)$. Second, we adapt the approach of adding additional localizing constraints from Section 3.2 to the commutative setting, where we do show $\xi_{\infty,S^{n-1}}^{\mathrm{cp}}(A) = \xi_{*,S^{n-1}}^{\mathrm{cp}}(A) = \tau_{\mathrm{cp}}(A)$. This yields a doubly indexed sequence of semidefinite programs whose optimal values converge to $\tau_{\mathrm{cp}}(A)$.

Proposition 4.5. *Let $A \in \mathcal{CP}^n$. For every $t \in \mathbb{N} \cup \{\infty\}$ the optimum in $\xi_{t,\dagger}^{\mathrm{cp}}(A)$ is attained. We have $\xi_{t,\dagger}^{\mathrm{cp}}(A) \to \xi_{\infty,\dagger}^{\mathrm{cp}}(A)$ and $\xi_{*,\dagger}^{\mathrm{cp}}(A) = \tau_{\mathrm{cp}}(A)$.*

Proof. The attainment of the optima in $\xi_{t,\dagger}^{\mathrm{cp}}(A)$ for $t \in \mathbb{N} \cup \{\infty\}$ and the convergence of $\xi_{t,\dagger}^{\mathrm{cp}}(A)$ to $\xi_{\infty,\dagger}^{\mathrm{cp}}(A)$ can be shown in the same way as the analogue statements for $\xi_t^{\mathrm{cp}}(A)$ in Proposition 4.3.

We have seen the inequality $\xi_{*,\dagger}^{\mathrm{cp}}(A) \leq \tau_{\mathrm{cp}}(A)$ in Proposition 4.1. Now we show the reverse inequality. Let L be feasible for $\xi_{*,\dagger}^{\mathrm{cp}}(A)$. We will show that L is feasible for $\tau_{\mathrm{cp}}(A)$, which implies $\tau_{\mathrm{cp}}(A) \leq L(1)$ and thus $\tau_{\mathrm{cp}}(A) \leq \xi_{*,\dagger}^{\mathrm{cp}}(A)$.

By Proposition 4.2 and the fact that $\mathrm{rank}(A) \leq \tau_{\mathrm{cp}}^{\mathrm{sos}}(A)$ we have $L(1) > 0$ (assuming $A \neq 0$). By Theorem 2.6, 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 $v_k \in D(S_A^{\mathrm{cp}})$. We define the matrices $R_k = v_k v_k^{\mathsf{T}}$, so that $A = L(1) \sum_{k=1}^{K} R_k$. The matrices R_k satisfy $0 \leq R_k \leq A$ since $v_k \in D(S_A^{\mathrm{cp}})$. Clearly also $R_k \succeq 0$. It remains to show that $R_k \preceq A$. For this we use the tensor constraints (17). 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 $v^{\mathsf{T}} R_k v > 0$ we have

$$L(1)\lambda_k \leq \left(\frac{v^{\mathsf{T}} A v}{v^{\mathsf{T}} R_k v} \right)^l \quad \text{for all} \quad l \in \mathbb{N}. \tag{18}$$

Suppose there is a k such that $R_k \not\preceq A$. Then there exists a vector v such that $v^{\mathsf{T}} R_k v > v^{\mathsf{T}} A v$. As $(v^{\mathsf{T}} A v)/(v^{\mathsf{T}} R_k v) < 1$, letting l tend to ∞ we obtain $L(1)\lambda_k = 0$, reaching a contradiction. It follows that $R_k \preceq A$ for all $k \in [K]$. $\square$

The second approach for reaching $\tau_{\mathrm{cp}}(A)$ is based on using the extra localizing constraints from Section 3.2. For $V \subseteq S^{n-1}$, define $\xi_{t,V}^{\mathrm{cp}}(A)$ by replacing the truncated quadratic module $\mathcal{M}_{2t}(S_A^{\mathrm{cp}})$ in $\xi_t^{\mathrm{cp}}(A)$ by $\mathcal{M}_{2t}(S_{A,V}^{\mathrm{cp}})$, where

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

Proposition 3.6 can be adapted to the completely positive setting, so that we have a sequence of finite subsets $V_1 \subseteq V_2 \subseteq \ldots \subseteq S^{n-1}$ with $\xi_{*,V_k}^{\mathrm{cp}}(A) \to \xi_{*,S^{n-1}}^{\mathrm{cp}}(A)$ as $k \to \infty$. Proposition 4.3 still holds when adding extra localizing constraints, so that for any $k \geq 1$ we have

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

Combining this with Proposition 4.6 below shows that we have a doubly indexed sequence $\xi_{t,V_k}^{\mathrm{cp}}(A)$ of (finite) semidefinite programs that converges to $\tau_{\mathrm{cp}}(A)$ as $t \to \infty$ and $k \to \infty$.

Proposition 4.6. *For $A \in \mathcal{CP}^n$ we have $\xi_{*,S^{n-1}}^{\mathrm{cp}}(A) = \tau_{\mathrm{cp}}(A)$.*

Proof. The proof is the same as the proof of Proposition 4.4, with the following additional observation: Given a vector $u \in \mathbb{R}^n$, we have $u \in D(S_{A,S^{n-1}}^{\mathrm{cp}})$ only if $uu^\mathsf{T} \preceq A$. The latter follows from the additional localizing constraints: for each $v \in \mathbb{R}^n$ we have

$$0 \le v^\mathsf{T} A v - \Big(\sum_i v_i u_i \Big)^2 = v^\mathsf{T}(A - uu^\mathsf{T})v. \qquad \square$$

4.4 More efficient tensor constraints

Here we show that for any integer $l \ge 2$ 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

$$(Q_l)_{m,w} = \begin{cases} 1/d_m & \text{if } w^c = m, \\ 0 & \text{otherwise,} \end{cases} \tag{19}$$

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

$$d_m = \big| \{ w \in \langle \mathbf{x} \rangle_{=l} : w^c = m \} \big| = \frac{l!}{\alpha_1! \cdots \alpha_n!}. \tag{20}$$

Lemma 4.7. *For $L \in \mathbb{R}[\mathbf{x}]_{2l}^*$ we have*

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

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

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

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

$$P = \frac{1}{l!} \sum_{\sigma \in S_l} P_\sigma, \tag{21}$$

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

$$M \succeq 0 \iff Q_l M Q_l^\mathsf{T} \succeq 0.$$

Proof. The implication $M \succeq 0 \implies Q_l M Q_l^\mathsf{T} \succeq 0$ 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 $Q_l^\mathsf{T} D Q_l = P$, the matrix in (21). Indeed, for any $w, w' \in \langle \mathbf{x} \rangle_{=l}$, we have

$$(Q_l^\mathsf{T} D Q_l)_{ww'} = \sum_{m \in [\mathbf{x}]_{=l}} (Q_l)_{mw} (Q_l)_{mw'} D_{mm} = \begin{cases} 1/d_m & \text{if } w^c = (w')^c = m, \\ 0 & \text{otherwise} \end{cases}$$

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

Suppose $Q_l M Q_l^\mathsf{T} \succeq 0$, and let λ 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 λ. We may also assume z to be a unit vector. Then $\lambda \geq 0$ can be shown using the identity $Q_l^\mathsf{T} D Q_l = P$ as follows:

$$\lambda = z^\mathsf{T} M z = z^\mathsf{T} P M P z = z^\mathsf{T} (Q_l^\mathsf{T} D Q_l) M (Q_l^\mathsf{T} D Q_l) z = (DQ_l z)^\mathsf{T} (Q_l M Q_l^\mathsf{T}) DQ_l z \geq 0. \qquad \square$$

We can now derive our symmetry reduction result:

Proposition 4.9. *For $L \in \mathbb{R}[\mathbf{x}]_{2l}^*$ we have*

$$A^{\otimes l} - (L((ww')^c))_{w,w' \in \langle \mathbf{x} \rangle_{=l}} \succeq 0 \quad \Longleftrightarrow \quad Q_l A^{\otimes l} Q_l^\mathsf{T} - (L(mm'))_{m,m' \in [\mathbf{x}]_{=l}} \succeq 0.$$

Proof. For any $w, w' \in \langle \mathbf{x} \rangle_{=l}$ we have $(P_\sigma A^{\otimes l} P_\sigma^\mathsf{T})_{w,w'} = A^{\otimes l}_{w^\sigma, (w')^\sigma} = A^{\otimes l}_{w,w'}$ 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 Lemma 4.7 and Lemma 4.8. $\qquad \square$

4.5 Computational examples

4.5.1 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 5×5 completely positive matrices with at least one zero entry have cp-rank at most 6; see [BSM03, Theorem 3.12]. Fawzi and Parrilo [FP16] show $\tau_{\mathrm{cp}}^{\mathrm{sos}}(P(0,0)) = 6$, 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 4.10. *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

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

It follows that c is the all ones vector $c = \mathbf{1}$. Moreover, if $P(a,b)_{ij} = 0$ for some $i \ne j$, then the constraints $L(x_i x_j u) \ge 0$ and $L((P(a,b)_{ij} - x_i x_j)u) \ge 0$ 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 $x_i x_j \ne x_k x_l$. 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

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

Finally, by the constraints $L((P(a,b)_{ij} - x_i x_j)u) \ge 0$ (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 $z_k \le 1$ for all $k \in [pq]$. Combined with the above inequality, it follows that

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

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

4.5.2 Examples related to the DJL-conjecture

The Drew-Johnson-Loewy conjecture [DJL94] 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 [BSU14]. 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 [BSU14], so that the diagonal entries become equal to 1. The matrices $\tilde{M}_7$, $\tilde{M}_8$ and $\tilde{M}_9$ correspond to the matrices $\tilde{M}$ in Examples 1,2,3 of [BSU14], and M_7, M_{11} correspond to the matrices M in Examples 1 and 4. The column $\xi_{2,\dagger}^{\mathrm{cp}}(\cdot) + x_i x_j$ corresponds to the bound $\xi_{2,\dagger}^{\mathrm{cp}}(\cdot)$ where we replace S_A^{cp} by $S_A^{\mathrm{cp}} \cup \{x_i x_j : 1 \le i < j \le n\}$.

Example	cp-rank$(\cdot)$	$\lfloor \frac{n^2}{4} \rfloor$	rank$(\cdot)$	$\xi_1^{\mathrm{cp}}(\cdot)$	$\xi_2^{\mathrm{cp}}(\cdot)$	$\xi_{2,\dagger}^{\mathrm{cp}}(\cdot)$	$\xi_{2,\dagger}^{\mathrm{cp}}(\cdot) + x_i x_j$	$\xi_{3,\dagger}^{\mathrm{cp}}(\cdot)$
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	–

Table 1: Examples from [BSU14] with various bounds on their cp-rank.

5 Lower bounds on the nonnegative rank

In this section we adapt the techniques for the cp-rank from Section 4 to the asymmetric setting of the nonnegative rank. We may 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 $\mathrm{rank}_+(A)$, 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 $\mathbf{X} \in \mathcal{D}(S_A^+)$.

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

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

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

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

$$L(gu) \geq 0 \quad \text{for} \quad g \in \{1\} \cup S_A^+ \quad \text{and} \quad u \in [\mathbf{x}]_{2t - \deg(g)}. \tag{22}$$

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

5.1 Comparison to other bounds

As in the previous section, we compare our bounds to the bounds by Fawzi and Parrilo [FP16]. They introduce the following parameter $\tau_+(A)$ as analogue of the bound $\tau_{\mathrm{cp}}(A)$ for the nonnegative rank:

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

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

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

First we give the analogue of Proposition 4.3, whose proof we omit since it is very similar.

Proposition 5.1. *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 $\xi_t^+(A) \to \xi_\infty^+(A) = \xi_*^+(A)$. 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 5.2. *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 Section 5 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 5.1 in terms of conic evaluations, after observing 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 $0 \leq R \leq A$. $\qquad \square$

Analogous to 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 5.3. *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 [FP16] summarizes how $\tau_+^{\mathrm{sos}}(A)$ and $\tau_+(A)$ relate to the combinatorial lower bounds

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

Here $\mathrm{RG}(A)$ is the *rectangular graph*, with vertex set $V = \{(i, j) \in [m] \times [n] : A_{ij} > 0\}$ and edge set $E = \{((i, j), (k, l)) : A_{il} A_{kj} = 0\}$. 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 [FMP$^+$15] 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., [FKPT13]). 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 [FP16] have furthermore shown that the bound $\tau_+(A)$ is at least as good as norm-based lower bounds:

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

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

$$\mathcal{N}^*(A) = \max\{L(A) : L : \mathbb{R}^{m \times n} \to \mathbb{R} \text{ linear and } L(X) \leq 1 \text{ for all } X \in \mathbb{R}_{\geq 0}^{m \times n}$$
$$\text{with } \mathrm{rank}(X) \leq 1 \text{ and } \mathcal{N}(X) \leq 1\}.$$

These bounds are called *norm-based* since norms often provide valid functions $\mathcal{N}$. For example, when $\mathcal{N}$ is the ℓ_∞-norm, Rothvoß [Rot14] has used the corresponding lower bound to show that the matching polytope has exponential extension complexity.

When $\mathcal{N}$ is the Frobenius norm: $\mathcal{N}(A) = \sqrt{\sum_{i,j} A_{ij}^2}$, the parameter $\mathcal{N}^*(A)$ is known as the *nonnegative nuclear norm*. It is denoted by $\nu_+(A)$ in [FP15], they show that it satisfies

$\text{rank}_+(A) \geq \left(\frac{\nu_+(A)}{||A||_F}\right)^2$, and that $\nu_+(A)$ admits the following reformulations:

$$\nu_+(A) = \min\Big\{\sum_i \lambda_i : A = \sum_i \lambda_i u_i v_i^\mathsf{T}, \ (\lambda_i, u_i, v_i) \in \mathbb{R}_+ \times \mathbb{R}_+^m \times \mathbb{R}_+^n, \ ||u_i||_2 = ||v_i||_2 = 1\Big\} \quad (23)$$

$$= \max\Big\{\langle A, W\rangle : W \in \mathbb{R}^{m \times n}, \ \big(\begin{smallmatrix} I & -W \\ -W^\mathsf{T} & I \end{smallmatrix}\big) \text{ is copositive}\Big\}, \quad (24)$$

where the cone of copositive matrices is the dual of the cone of completely positive matrices. Fawzi and Parrilo [FP15] use the copositive formulation (24) to provide bounds $\nu_+^{[k]}(A)$ ($k \geq 0$), based on inner approximations of the copositive cone from [Par00], which converge to $\nu_+(A)$ from below. We now observe that by Theorem 2.8 the atomic formulation of $\nu_+(A)$ from (23) can be seen as a moment optimization problem:

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

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

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

(The same observation is made in [TS15] for the real nuclear norm of a symmetric 3-tensor and in [Nie16] 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 $L \geq 0$ on $\mathcal{M}_{2t}(S_A^+)$ by $L \geq 0$ on $\mathcal{M}_{2t}(\{x_1, \ldots, x_{m+n}\})$ and $L = 0$ on $\mathcal{I}_{2t}(S)$, and let $\mu_*(A)$ be obtained by adding the constraint $\text{rank}(M(L)) < \infty$ to $\mu_\infty(A)$. We have $\mu_t(A) \to \mu_\infty(A) = \mu_*(A) = \nu_+(A)$ by Theorem 2.8 and (a non-normalized analogue of) Theorem 2.9. One can show that $\mu_1(A)$ with the additional constraints $L(u) \geq 0$ 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.

5.2 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 τ_+ and τ_+^{sos} on the two examples considered in [FP16].

5.2.1 All nonnegative 2×2 matrices

For $A(\alpha) = \big(\begin{smallmatrix} 1 & 1 \\ 1 & \alpha \end{smallmatrix}\big)$, Fawzi and Parrilo [FP16] show that

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

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 2×2 matrices. By using a semidefinite programming solver for $\alpha = k/100$, $k \in [100]$, we see that $\xi_2^+(A(\alpha))$ coincides with $\tau_+(A(\alpha))$.

5.2.2 The nested rectangles problem

In this section we consider the nested rectangles problem as described in [FP16, Section 2.7.2] (see also [MSvS03]), 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 [Yan91], but also to extended formulations of nested pairs [BFPS15, GG12]. 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. 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. [BFPS15, Theorem 1]) that the extension complexity of the pair (P_1, P_2), where

$$P_1 = \text{conv}(\{v_1, \ldots, v_n\}) \quad \text{and} \quad P_2 = \{x : a_i^\mathsf{T} x \le b_i \text{ for } i \in [m]\},$$

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 [GPT13, Lemma 4.1] (see also [GG12]).

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 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 $(1 + a)(1 + b) \le 2$ (see [FP16, Proposition 4] for a proof sketch). To test the quality of their bound, Fawzi and Parrilo [FP16] compute $\tau_+^{\text{sos}}(S(a, b))$ for different values of a and b. In doing so they determine the region where $\tau_+^{\text{sos}}(S(a, b)) > 3$. 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 Figure 1. The results show that $\xi_{2,\dagger}^+(S(a, b))$ strictly improves upon the bound $\tau_+^{\text{sos}}(S(a, b))$, and that $\xi_{3,\dagger}^+(S(a, b))$ is again a strict improvement over $\xi_{2,\dagger}^+(S(a, b))$.

6 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 [LWdW17, Lemma 5] to rescale a factorization and claim that, for any $A \in \mathbb{R}^{m \times n}_+$ with psd-rank$_\mathbb{C}(A) = d$, there exists a factorization $A = (\langle X_i, X_{m+j} \rangle)$ by matrices $X_1, \ldots, X_{m+n} \in \mathcal{H}^d_+$ such that $\sum_{i=1}^m X_i = I$ and $\text{Tr}(X_{m+j}) = \sum_i A_{ij}$ for all $j \in [n]$. Indeed, starting from any factorization $X_i, X_{m+j} \in \mathcal{H}^d_+$ 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

$$S_A^{\text{psd}} = \left\{ x_i - x_i^2 : i \in [m] \right\} \cup \left\{ \left(\sum_{i=1}^m A_{ij} \right) x_{m+j} - x_{m+j}^2 : j \in [n] \right\}$$
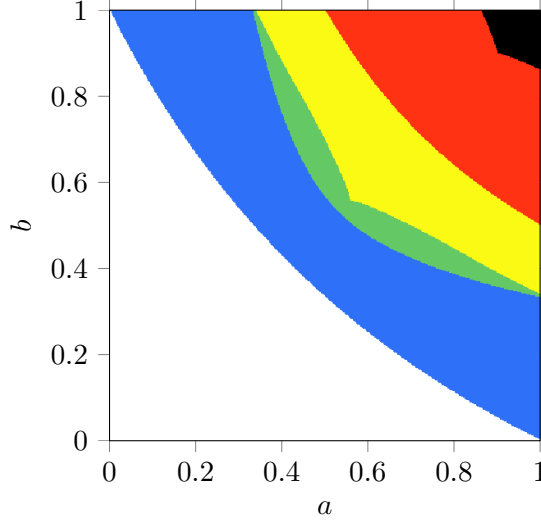
Figure 1: The colored region corresponds to $\mathrm{rank}_+(S(a,b)) = 4$. The top right region (black) corresponds to $\xi_{1,\dagger}^+(S(a,b)) > 3$, the two top right regions (black and red) together correspond to $\tau_+^{\mathrm{sos}}(S(a,b)) > 3$, the three top right regions (black, red and yellow) to $\xi_{2,\dagger}^+(S(a,b)) > 3$, and the four top right regions (black, red, yellow, and green) to $\xi_{3,\dagger}^+(S(a,b)) > 3$.

is localizing for A; that is, there is *at least one* minimal factorization $\mathbf{X}$ of A such that $g(\mathbf{X}) \succeq 0$ for all polynomials $g \in S_A^{\mathrm{psd}}$. Moreover, for the same minimal factorization $\mathbf{X}$ 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 $A \in \mathbb{R}_{\geq 0}^{m \times n}$, 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 $\xi_\infty^{\mathrm{psd}}(A)$. By the above discussion it follows that the parameter $\xi_*^{\mathrm{psd}}(A)$ is a lower bound on $\mathrm{psd\text{-}rank}_{\mathbb{C}}(A)$ and we have

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

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 the examples in Section 6.2.2).

It follows from the construction of S_A^{psd} and Equation (7) that the quadratic module $\mathcal{M}(S_A^{\mathrm{psd}})$ is Archimedean, and hence the following analogue of Proposition 3.1 can be shown.

Proposition 6.1. *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 smallest $\alpha \geq 0$ for which there exists a unital C^*-algebra $\mathcal{A}$ with tracial state τ 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 $A = \alpha \cdot (\tau(X_i X_{m+j}))_{i \in [m], j \in [n]}$.

6.1 Comparison to other bounds

In [LWdW17] the following bound on the complex positive semidefinite rank was derived:

$$\text{psd-rank}_{\mathbb{C}}(A) \geq \sum_{i=1}^{m} \max_{j \in [n]} \frac{A_{ij}}{\sum_i A_{ij}}. \tag{25}$$

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

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

We observe that the inequalities $L(x_i(\sum_i A_{ij} - x_{m+j})) \geq 0$ are of the form used in Lemma 3.9 and are indeed valid strengthenings of $\xi_t^{\text{psd}}(A)$. More importantly, as in Lemma 3.9, 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 (25).

In [LWdW17] two other fidelity based lower bounds on the psd-rank were defined; we do not know how they compare to $\xi_t^{\text{psd}}(A)$.

6.2 Computational examples

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

6.2.1 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 $\text{psd-rank}_{\mathbb{R}}(M(b,c)) \geq \text{psd-rank}_{\mathbb{C}}(M(b,c)) \geq 2$. In [FGP+15, Example 2.7] it is shown that

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

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

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 Figure 2 we see that the bound $\xi_2^{\text{psd}}(M(b,c))$ permits to certify $\text{psd-rank}_{\mathbb{R}}(M(b,c)) = \text{psd-rank}_{\mathbb{C}}(M(b,c)) = 3$ for most values (b,c) where $\text{psd-rank}_{\mathbb{R}}(M(b,c)) = 3$.

6.2.2 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.
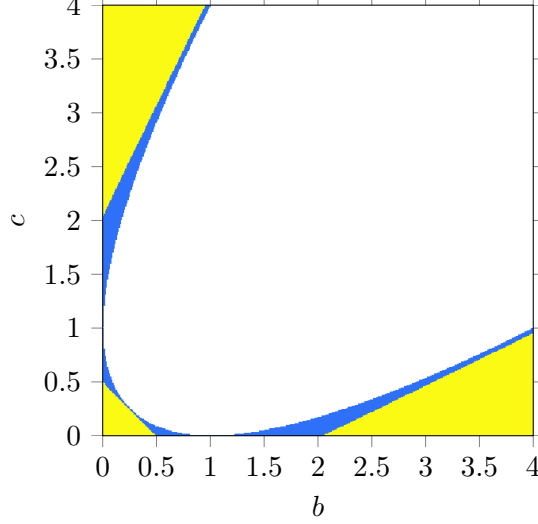
Figure 2: The colored region corresponds to the values (b, c) for which $\text{psd-rank}_{\mathbb{R}}(M(b,c)) = 3$; the outer region (yellow) shows the values of (b, c) for which $\xi_2^{\text{psd}}(M(b,c)) > 2$.

We consider the quadrilateral Q whose vertices are $(0,0), (0,1), (1,0), (2,2)$, 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 $\text{psd-rank}_{\mathbb{R}}(S_Q) = 3$ (a corollary of [GRT13, Theorem 4.3]) and therefore both bounds certify $\text{psd-rank}_{\mathbb{C}}(S_Q) = 3 = \text{psd-rank}_{\mathbb{R}}(S_Q)$.

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 [GGS16] it is shown that $\text{psd-rank}_{\mathbb{C}}(S_H) = 3$.

7 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: 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 Section 5 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 $\xi_t^+(A)$ 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 $i_1 \in [n_1], i_2 \in [n_2], i_3 \in [n_3]$), and 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 $\tau_+^{\text{sos}}(A)$ from [FP16]. One can show $\xi_*^+(A) = \tau_+(A)$,

and, by 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 $\xi_3^+(A)$, one can show $\xi_{3,\dagger}^+(A) \geq \tau_+^{\mathrm{sos}}(A)$.

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 4.3 that if A is not completely positive then $\xi_t^{\mathrm{cp}}(A)$ is infeasible for all t large enough (which is analogous to an earlier result in [Nie14]). In the noncommutative setting the situation is more complicated: If $\xi_*^{\mathrm{cpsd}}(A)$ is feasible, then $A \in \mathcal{CS}_+^n$, and if $A \notin \mathcal{CS}_{\mathrm{vN+}}^n$, then $\xi_t^{\mathrm{cpsd}}(A)$ is infeasible for all t large enough (Propositions 3.1 and 3.2). Here $\mathcal{CS}_{\mathrm{vN+}}$ is the cone defined in [BLP17] consisting of the matrices admitting a factorization in a von Neumann algebra with a trace. By Lemma 2.10, $\mathcal{CS}_{\mathrm{vN+}}$ can equivalently be characterized as the set of matrices of the form $\alpha\left(\tau(a_i a_j)\right)$ for some C^*-algebra $\mathcal{A}$ with tracial state τ, positive elements $a_1, \ldots, a_n \in \mathcal{A}$ 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 [PSVW16, GdLL17a, PV17], 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 5×5 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 [Rot14] (indirectly) shows that the parameter ξ_∞^+ 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 ξ_∞^+, 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)?

Acknowledgments

A   Proofs

Proofs for Section 2.1

Theorem 2.1. *Let $S \subseteq \mathrm{Sym}\,\mathbb{R}\langle \mathbf{x}\rangle$ 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 $\mathcal{A}$ with tracial state τ and $\mathbf{X} \in \mathcal{D}_\mathcal{A}(S) \cap \mathcal{V}_\mathcal{A}(T)$ with

$$L(p) = \tau(p(\mathbf{X})) \quad \text{for all} \quad p \in \mathbb{R}\langle \mathbf{x}\rangle. \tag{3}$$

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

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

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

Now 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 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

$$X_i \colon \mathbb{R}\langle \mathbf{x} \rangle / N(L) \to \mathbb{R}\langle \mathbf{x} \rangle / N(L), \ \overline{p} \mapsto \overline{x_i p} \tag{26}$$

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 $R - \sum_{i=1}^n x_i^2 \in \mathcal{M}(S) + \mathcal{I}(T)$. Using $R - x_i^2 = (R - \sum_{j=1}^n x_j^2) + \sum_{j \neq i} x_j^2 \in \mathcal{M}(S) + \mathcal{I}(T)$ 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 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 $f \in \mathbb{R}\langle \mathbf{x} \rangle$ and $h \in T$.

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

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

$$\tau(p(\mathbf{X}) q(\mathbf{X})) = \langle \overline{1}, p(\mathbf{X}) q(\mathbf{X}) \overline{1} \rangle = \langle \overline{1}, \overline{pq} \rangle = L(pq), \tag{27}$$

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

To conclude the proof, observe that (3) follows from (27) by taking $q = 1$. $\qquad \square$

Theorem 2.2. *For $S \subseteq \mathrm{Sym}\,\mathbb{R}\langle \mathbf{x} \rangle$, $T \subseteq \mathbb{R}\langle \mathbf{x} \rangle$, 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 $\mathcal{M}(S)$, zero on $\mathcal{I}(T)$, and has $\mathrm{rank}(M(L)) < \infty$;*

(2) *there is a finite dimensional C^*-algebra $\mathcal{A}$, a tracial state τ, and $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S) \cap \mathcal{V}_{\mathcal{A}}(T)$ satisfying equation (3);*

(3) *L is a convex combination of normalized trace evaluations at points in $\mathcal{D}(S) \cap \mathcal{V}(T)$.*

Proof. ((1) $\Rightarrow$ (2)) Here we can follow the proof of Theorem 2.1, 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 $\mathcal{A}$ is finite dimensional.

((2) $\Rightarrow$ (3)) By Artin-Wedderburn theory we have the $*$-isomorphism

$$\varphi \colon \mathcal{A} \to \bigoplus_{m=1}^M \mathbb{C}^{d_m \times d_m}.$$

Define the $*$-homomorphisms $\varphi_m \colon \mathcal{A} \to \mathbb{C}^{d_m \times d_m}$ by $\varphi = \oplus_m \varphi_m$. 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 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 $\tau(a) = \sum_m \lambda_m \, \mathrm{tr}(\varphi_m(a))$ for nonnegative scalars λ_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 φ_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 $(X_1^m, \ldots, X_n^m) \in \mathcal{D}(S) \cap \mathcal{V}(T)$.

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

Theorem 2.3. *Let $1 \le \delta \le t < \infty$, $S \subseteq \mathrm{Sym}\,\mathbb{R}\langle \mathbf{x} \rangle_{2\delta}$, and $T \subseteq \mathbb{R}\langle \mathbf{x} \rangle_{2\delta}$. Suppose $L \in \mathbb{R}\langle \mathbf{x} \rangle_{2t}^*$ is symmetric, tracial, δ-flat, nonnegative on $\mathcal{M}_{2t}(S)$, and zero on $\mathcal{I}_{2t}(T)$. Then L extends to a symmetric, tracial, linear form on $\mathbb{R}\langle \mathbf{x} \rangle$ that is nonnegative on $\mathcal{M}(S)$, zero on $\mathcal{I}(T)$, 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

$$\mathbb{R}\langle \mathbf{x} \rangle_t = \mathrm{span}(W) \oplus N_t(L). \tag{28}$$

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 $\deg(p) \le 2t$, 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 $\deg(x_i p) \le t$, 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

$$\hat{L}((x_i u)^* v x_j) = \hat{L}((x_i u)^* r_v x_j) = \hat{L}((r_v x_j)^* x_i u) = \hat{L}((r_v x_j)^* x_i r_u) = \hat{L}((x_i r_u)^* r_v x_j), \tag{29}$$

and in an analogous way one can show

$$\hat{L}((u x_i)^* x_j v) = \hat{L}((r_u x_i)^* x_j r_v). \tag{30}$$

We can now show that $\hat{L}$ is tracial. We do this by showing that $\hat{L}(w x_j) = \hat{L}(x_j w)$ for all w with $\deg(w) \le 2t + 1$. Notice that when $\deg(w) \le 2t - 1$ the statement follows from the fact that $\hat{L}$ is an extension of L. Suppose $w = u^* v$ with $\deg(u) = t + 1$ and $\deg(v) \le t$. 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

$$
\begin{aligned}
\hat{L}(w x_j) = \hat{L}(u^* v x_j) &= \hat{L}((x_i u')^* v x_j) \\
&= \hat{L}((x_i r_{u'})^* r_v x_j) && \text{by (29)} \\
&= L((x_i r_{u'})^* r_v x_j) && \text{since } \deg(x_i r_{u'} r_v x_j) \le 2t \\
&= L((r_{u'} x_j)^* x_i r_v) && \text{since } L \text{ is tracial} \\
&= \hat{L}((r_{u'} x_j)^* x_i r_v) && \text{since } \deg((r_{u'} x_j)^* x_i r_v) \le 2t \\
&= \hat{L}((u' x_j)^* x_i v) && \text{by (30)} \\
&= \hat{L}(x_j w).
\end{aligned}
$$

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 $\mathcal{M}(S)$ and zero on $\mathcal{I}(T)$. For this we make two observations:

(i) $\mathcal{I}(N_t(L)) \subseteq N(\hat{L})$.

(ii) $\mathbb{R}\langle \mathbf{x} \rangle = \operatorname{span}(W) \oplus \mathcal{I}(N_t(L))$.

For (i) we use the (easy to check) fact that $N_t(L) = \operatorname{span}(\{u - r_u : u \in \langle \mathbf{x} \rangle_t\})$. 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 $\operatorname{span}(W) \cap N(\hat{L}) = \{0\}$, so we have the direct sum $\operatorname{span}(W) \oplus \mathcal{I}(N_t(L))$. 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 (28). Let $w = x_i v \in \langle \mathbf{x} \rangle$ and assume $v \in \operatorname{span}(W) \oplus \mathcal{I}(N_t(L))$, that is, $v = r_v + q_v$ where $r_v \in \operatorname{span}(W)$ and $q_v \in \mathcal{I}(N_t(L))$. We have $x_i v = x_i r_v + x_i q_v$ so it suffices to show $x_i r_v, x_i q_v \in \operatorname{span}(W) \oplus \mathcal{I}(N_t(L))$. Clearly $x_i q_v \in \mathcal{I}(N_t(L))$, since $q_v \in \mathcal{I}(N_t(L))$. Also, observe that $x_i r_v \in \mathbb{R}\langle \mathbf{x} \rangle_t$ and therefore $x_i r_v \in \operatorname{span}(W) \oplus \mathcal{I}(N_t(L))$ by (28).

We conclude the proof by showing that $\hat{L}$ is nonnegative on $\mathcal{M}(S)$ and zero on $\mathcal{I}(T)$. Let $g \in \mathcal{M}(S)$, $h \in \mathcal{I}(T)$, 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 $r_p \in \operatorname{span}(W)$ and $p - r_p \in \mathcal{I}(N_t(L))$, 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 $\deg(r_p h) \leq (t - \delta) + 2\delta \leq 2t$. $\qquad\square$

Proofs for Section 2.2

Theorem 2.5. *Let $S, T \subseteq \mathbb{R}[\mathbf{x}]$ with $\mathcal{M}(S) + \mathcal{I}(T)$ Archimedean. For $L \in \mathbb{R}[\mathbf{x}]^*$, the following are equivalent:*

(1) *L is nonnegative on $\mathcal{M}(S)$, zero on $\mathcal{I}(T)$, and $L(1) = 1$;*

(2) *there exists a unital commutative C^*-algebra $\mathcal{A}$ with a state τ and $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S) \cap \mathcal{V}_{\mathcal{A}}(T)$ such that $L(p) = \tau(p(\mathbf{X}))$ for all $p \in \mathbb{R}[\mathbf{x}]$;*

(3) *there is a probability measure μ 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 analogue of the implication $(1) \Rightarrow (2)$ in Theorem 2.1 (observing in addition that the operators X_i in (26) pairwise commute so that the constructed C^*-algebra $\mathcal{A}$ is commutative).

$((2) \Rightarrow (3))$ Let $\widehat{\mathcal{A}}$ denote the set of unital $*$-homomorphisms $\mathcal{A} \to \mathbb{C}$, known as the *spectrum* of $\mathcal{A}$. We equip $\widehat{\mathcal{A}}$ with the weak-$*$ topology, so that it is compact as a result of $\mathcal{A}$ being unital (see, e.g., [Bla06, II.2.1.4]). The *Gelfand representation* is the $*$-isomorphism

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

where $\mathcal{C}(\widehat{\mathcal{A}})$ is the set of complex-valued continuous functions on $\widehat{\mathcal{A}}$. Since Γ is an isomorphism, the state τ on $\mathcal{A}$ induces a state τ' on $\mathcal{C}(\widehat{\mathcal{A}})$ defined by $\tau'(\Gamma(a)) = \tau(a)$ for $a \in \mathcal{A}$. By the Riesz

representation theorem (see, e.g., [Rud87, Theorem 2.14]) there is a Radon measure ν on $\widehat{\mathcal{A}}$ such that

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

We then have

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

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

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

Since $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S)$, we have $g(\mathbf{X}) \succeq 0$ for all $g \in S$, hence $\Gamma(g(\mathbf{X}))$ is a positive element of $\mathcal{C}(\widehat{\mathcal{A}})$, 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 $D(S) \cap V(T)$, μ is a probability measure on $D(S) \cap V(T)$ since $L(1) = 1$, and we have

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

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

Note that the more common proof for the implication $(1) \Rightarrow (3)$ in Theorem 2.5 relies on Putinar's Positivstellensatz [Put93]: if L satisfies (1) then $L(p) \geq 0$ for all polynomials p nonnegative on $D(S) \cap V(T)$ (since $p + \varepsilon \in \mathcal{M}(S) + \mathcal{I}(T)$ for any $\varepsilon > 0$), and thus L has a representing measure μ as in (3) by the Riesz-Haviland theorem [Hav36].

Theorem 2.6. *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 $\mathcal{M}(S)$, zero on $\mathcal{I}(T)$, has $\operatorname{rank}(M(L)) < \infty$, and $L(1) = 1$;

(2) there is a finite dimensional commutative C^*-algebra $\mathcal{A}$, a state τ, and $\mathbf{X} \in \mathcal{D}_{\mathcal{A}}(S) \cap \mathcal{V}_{\mathcal{A}}(T)$ satisfying equation (3);

(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 analogue. 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

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

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 $\mathcal{M}(S')$, zero on $\mathcal{I}(T')$, and satisfies $\operatorname{rank} M(L') = \operatorname{rank} M(L) < \infty$. Following the proof of the implication $(1) \Rightarrow (2)$ in Theorem 2.1, we see that the operators $X_1, \ldots, X_n$ pairwise commute (since $\mathbf{X} \in \mathcal{V}_{\mathcal{A}}(T')$ and T' contains all $x_i x_j - x_j x_i$) and thus the constructed C^*-algebra $\mathcal{A}$ is finite dimensional and commutative.

$((2) \Rightarrow (3))$ Here we follow the proof of this implication in Theorem 2.2 and observe that since $\mathcal{A}$ 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. $\qquad \square$

Theorem 2.7. *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 δ-flat, nonnegative on $\mathcal{M}_{2t}(S)$, and zero on $\mathcal{I}_{2t}(T)$, 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 analogue in Corollary 2.4. As in the above proof for the implication (1) $\implies$ (2) in Theorem 2.6, 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 $\mathcal{M}_{2t}(S')$, zero on $\mathcal{I}_{2t}(T')$, and δ-flat. By Corollary 2.4, L' is a conic combination of trace evaluation maps at elements of $\mathcal{D}(S') \cap \mathcal{V}(T')$. 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 $\mathbf{X} \in \mathcal{V}(T')$, the matrices $X_1, \ldots, X_n$ pairwise commute and thus can be assumed to be diagonal. Since $\mathbf{X} \in \mathcal{D}(S') \cap \mathcal{V}(T')$, 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 $r_j = ((X_1)_{jj}, \ldots, (X_n)_{jj}) \in D(S) \cap V(T)$. $\qquad\square$

Theorem 2.8. *Let* $S, T \subseteq \mathbb{R}[\mathbf{x}]$ *such that* $\mathcal{M}(S) + \mathcal{I}(T)$ *is Archimedean. If* $L \in \mathbb{R}[\mathbf{x}]^*$ *is nonnegative on* $\mathcal{M}(S)$ *and zero on* $\mathcal{I}(T)$, *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 2.5 there exists a probability measure μ on $D(S)$ such that

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

A general version of Tchakaloff's theorem, as explained in [BT06], shows 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

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

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

Proofs for Section 2.3

We start with the next lemma, which we then use for the proof of Theorem 2.11 (the noncommutative case), from which we then derive Theorem 2.9 (the commutative case).

Lemma 2.12. *Let* $S \subseteq \operatorname{Sym} \mathbb{R}\langle\mathbf{x}\rangle$ *and* $R > 0$ *such that* $R - (x_1^2 + \cdots + x_n^2) \in \mathcal{M}_2(S)$. *If* $L_t \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}^*$ *is nonnegative on* $\mathcal{M}_{2t}(S)$, *then* $|L_t(w)| \leq R^{|w|/2} L_t(1)$ *for all* $w \in \langle\mathbf{x}\rangle_{2t}$. *If in addition* $\sup_t L_t(1) < \infty$, *then* (L_t) *has a pointwise converging subsequence in* $\mathbb{R}\langle\mathbf{x}\rangle^*$.

Proof. We first use induction on $0 \leq |w| \leq t$ to show that $L_t(w^*w) \leq R^{|w|} L_t(1)$ for all $w \in \langle\mathbf{x}\rangle_t$. Indeed, if $L_t(w^*w) \leq R^{|w|} L_t(1)$, then

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

where we use that $L_t(w^*(x_i^2 - R)w) \leq 0$ since $w^*(R - x_i^2)w \in \mathcal{M}_{2t}(S)$. Next we write any $w \in \langle\mathbf{x}\rangle_{2t}$ as $w = w_1^* w_2$ with $w_1, w_2 \in \langle\mathbf{x}\rangle_t$ 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 $c := \sup_t L_t(1) < \infty$. Then, for each t, extending $L_t \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}^*$ to $\mathbb{R}\langle\mathbf{x}\rangle^*$ by setting $L_t(w) = 0$ for words $|w| > 2t$, we have that the vector $(L_t(w)/(R^{|w|/2}c))_{w \in \langle\mathbf{x}\rangle}$ lies in the supremum norm unit ball of $\mathbb{R}^{\langle\mathbf{x}\rangle}$, which is compact by the Banach–Alaoglu theorem. It follows that the sequence (L_t) has a converging subsequence. $\qquad\square$

Theorem 2.11. *Let $1 \leq \delta \leq t < \infty$ and $S \cup \{f\} \subseteq \mathrm{Sym}\,\mathbb{R}\langle\mathbf{x}\rangle_{2\delta}$ with $\mathcal{D}(S) \neq \emptyset$.*

    (i) If $\mathcal{M}(S)$ is Archimedean, then $f_t^{tr} \to f_\infty^{\mathrm{tr}}$ as $t \to \infty$, and the optimal values in f_∞^{tr} and $f_{\mathrm{II}_1}^{\mathrm{tr}}$ are attained and equal.

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

Proof. We first show (i). As $\mathcal{M}(S)$ is Archimedean, $R - \sum_{i=1}^n x_i^2 \in \mathcal{M}(S)$ for some $R > 0$. To avoid technical details and simplify notation we will make the stronger assumption that the polynomial $R - \sum_{i=1}^n x_i^2$ belongs to $\mathcal{M}_2(S)$, see, e.g., [PNA10] for the general case.

    Since the bounds f_t^{tr} are monotone nondecreasing in t and upper bounded by f_∞^{tr}, the limit $\lim_{t\to\infty} f_t^{\mathrm{tr}}$ exists and it is at most f_∞^{tr}. If L is feasible for f_t^{tr}, then by Lemma 2.12 we have $|L(w)| \leq R^{|w|/2}$ for all $w \in \langle\mathbf{x}\rangle_{2t}$. Hence the feasible region of f_t^{tr} is compact and, since it is non-empty, there exists an optimal solution $L_t \in \mathbb{R}\langle\mathbf{x}\rangle_{2t}^*$ to the program defining f_t^{tr}.

    By Lemma 2.12, the sequence (L_t) has a pointwise converging subsequence. Let $L \in \mathbb{R}\langle\mathbf{x}\rangle^*$ be the pointwise limit. One can easily check that L is feasible for f_∞^{tr}. Hence L is optimal for f_∞^{tr} and thus $f_t^{\mathrm{tr}} = L_t(f) \to L(f) = f_\infty^{\mathrm{tr}}$.

    Next, since L is symmetric, tracial, and nonnegative on $\mathcal{M}(S)$, we can apply Theorem 2.1 to obtain a feasible solution $(\mathcal{A}, \tau, \mathbf{X})$ to $f_{\mathrm{II}_1}^{\mathrm{tr}}$ satisfying (3) 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_∞^{tr} and $f_{\mathrm{II}_1}^{\mathrm{tr}}$.

    Finally, part (ii) is derived as follows. If L is an optimal solution of f_t^{tr} that is δ-flat, then, by Corollary 2.4, it has an extension $\hat{L} \in \mathbb{R}\langle\mathbf{x}\rangle^*$ that is a conic combination of trace evaluations at elements of $\mathcal{D}(S)$. This shows $f_*^{\mathrm{tr}} \leq \hat{L}(f) = L(f)$, and thus the chain of equalities $f_t^{\mathrm{tr}} = f_\infty^{\mathrm{tr}} = f_*^{\mathrm{tr}} = f_{\mathrm{II}_1}^{\mathrm{tr}}$ holds. $\qquad\square$

    We now derive the commutative analogue of the above theorem.

Theorem 2.9. *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 $\mathcal{M}(S)$ is Archimedean, then $f_t \to f_\infty$ as $t \to \infty$, and the optimal values in f_∞ and f_* are attained and $f_\infty = f_*$.

    (ii) If f_t admits an optimal solution L that is δ-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 2.11 in the commutative setting we see that $f_t \to f_\infty$ and that the optimum is attained in f_∞. Let L be optimal for f_∞ and let k be greater than $\deg(f)$ and $\deg(g)$ for $g \in S$. By Theorem 2.8, 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 2.11(ii), where, instead of using Corollary 2.4, we now use its commutative analogue, Theorem 2.7. $\qquad\square$

References

[AL12]   M. F. Anjos and J. B. Lasserre. *Handbook on Semidefinite, Conic and Polynomial Optimization.* International Series in Operations Research & Management Science Series. Springer, 2012.

[AMR$^+$16]   A. Atserias, L. Mančinska, D.E. Roberson, R. Šámal, S. Severini, and A. Varvitsiotis. Quantum and non-signalling graph isomorphisms. *arXiv:1611.09837*, 2016.

[ApS17]   MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.0.0.81*, 2017.

[BCKP13]   S. Burgdorf, K. Cafuta, I. Klep, and J. Povh. The tracial moment problem and trace-optimization of polynomials. *Mathematical Programming*, 137(1):557–578, 2013.

[BEK78]   G. P. Barker, L. Q. Eifler, and T. P. Kezlan. A non-commutative spectral theorem. *Linear Algebra and its Applications*, 20(2):95–100, 1978.

[BEKS17]   J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1):65–98, 2017.

[BFPS15]   G. Braun, S. Fiorini, S. Pokutta, and D. Steurer. Approximation limits of linear programs (beyond hierarchies). *Mathematics of Operations Research*, 40(3):756–772, 2015. Appeared earlier in FOCS'12.

[BFS16]   M. Berta, O. Fawzi, and V. B. Scholz. Quantum bilinear optimization. *SIAM Journal on Optimization*, 26(3):1529–1564, 2016.

[BK12]   S. Burgdorf and I. Klep. The truncated tracial moment problem. *Journal of Operator Theory*, 68(1):141–163, 2012.

[BKP16]   S. Burgdorf, I. Klep, and J. Povh. *Optimization of Polynomials in Non-Commutative Variables*. Springer Briefs in Mathematics. Springer, 2016.

[Bla06]   B. Blackadar. *Operator Algebras: Theory of C*-Algebras and Von Neumann Algebras*. Encyclopaedia of Mathematical Sciences. Springer, 2006.

[BLP17]   S. Burgdorf, M. Laurent, and T. Piovesan. On the closure of the completely positive semidefinite cone and linear approximations to quantum colorings. *Electronic Journal of Linear Algebra*, 32:15–40, 2017.

[BR06]   A. Berman and U. G. Rothblum. A note on the computation of the cp-rank. *Linear Algebra and its Applications*, 419:1–7, 2006.

[BSM03]   A. Berman and N. Shaked-Monderer. *Completely Positive Matrices*. World Scienfic, 2003.

[BSU14]   I. M. Bomze, W. Schachinger, and R. Ullrich. From seven to eleven: Completely positive matrices with high cp-rank. *Linear Algebra and its Applications*, 459:208 – 221, 2014.

[BT06]   C. Bayer and J. Teichmann. The proof of Tchakaloff's theorem. *Proceedings of the American Mathematical Society*, 134:3035–3040, 2006.

[Bur09]   S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120(2):479–495, 2009.

[CCZ10]   M. Conforti, G. Cornuéjols, and G. Zambelli. Extended formulations in combinatorial optimization. *4OR*, 8:1–48, 2010.

[CF96]   R. E. Curto and L. A. Fialkow. *Solution of the Truncated Complex Moment Problem for Flat Data*. Memoirs of the American Mathematical Society. American Mathematical Society, 1996.

[DD12]   P. Dickinson and M. Dür. Linear-time complete positivity detection and decomposition of sparse matrices. *SIAM Journal on Matrix Analysis and Applications*, 33(3):701–720, 2012.

[DJL94]   J. H. Drew, C. R. Johnson, and R. Loewy. Completely positive matrices associated with M-matrices. *Linear and Multilinear Algebra*, 37(4):303–310, 1994.

[dKP02]   E. de Klerk and D. V. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM Journal on Optimization*, 12(4):875–892, 2002.

[Edm65]   J. Edmonds. Maximum matching and a polyhedron with $0, 1$ vertices. *Journal of Research of the National Bureau of Standards*, 69 B:125–130, 1965.

[FFGT15]   Y. Faenza, S. Fiorini, R. Grappe, and H.R. Tiwari. Extended formulations, non-negative factorizations and randomized communication protocols. *Mathematical Programming*, 153(1):75–94, 2015.

[FGP+15]   H. Fawzi, J. Gouveia, P. A. Parrilo, R. Z. Robinson, and R. R. Thomas. Positive semidefinite rank. *Mathematical Programming*, 153(1):133–177, 2015.

[FKPT13]   S. Fiorini, V. Kaibel, K. Pashkovich, and D.O. Theis. Combinatorial bounds on nonnegative rank and extended formulations. *Discrete Mathematics*, 313(1):67–83, 2013.

[FMP+15]   S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, and R. de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *Journal of the ACM*, 62(2):17:1–17:23, 2015. Appeared earlier in STOC'12.

[FP15]   H. Fawzi and P. A. Parrilo. Lower bounds on nonnegative rank via nonnegative nuclear norms. *Mathematical Programming*, 153(1):41–66, 2015.

[FP16]   H. Fawzi and P. A. Parrilo. Self-scaled bounds for atomic cone ranks: applications to nonnegative rank and cp-rank. *Mathematical Programming*, 158(1):417–465, 2016.

[GB14]   M. Grant and S. Boyd. CVX: Matlab Software for Disciplined Convex Programming, version 2.1. `http://cvxr.com/cvx`, 2014.

[GdLL17a] S. Gribling, D. de Laat, and M. Laurent. Matrices with high completely positive semidefinite rank. *Linear Algebra and its Applications*, 513:122 – 148, 2017.

[GdLL17b] S. Gribling, D. de Laat, and M. Laurent. Bounds on entanglement dimensions and quantum graph parameters via noncommutative polynomial optimization. In preparation, 2017.

[GG12] N. Gillis and F. Glineur. On the geometric interpretation of the nonnegative rank. *Linear Algebra and its Applications*, 437(11):2685 – 2712, 2012.

[GGS16] A. P. Goucha, J. Gouveia, and P. M. Silva. On ranks of regular polygons. *arXiv:1610.09868*, 2016.

[Gil17] N. Gillis. Introduction to nonnegative matrix factorization. *SIAG/OPT Views and News*, 25(1):7–16, 2017.

[Goe15] M. Goemans. Smallest compact formulation for the permutahedron. *Mathematical Programming*, 153(1):5–11, 2015.

[GPT13] J. Gouveia, P. A. Parrilo, and R. R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38(2):248–264, 2013.

[GRT13] J. Gouveia, R. Z. Robinson, and R. R. Thomas. Polytopes of minimum positive semidefinite rank. *Discrete & Computational Geometry*, 50(3):679–699, 2013.

[Hav36] E. K. Haviland. On the Momentum Problem for Distribution Functions in More Than One Dimension. II. *American Journal of Mathematics*, 58(1):164–168, 1936.

[JSWZ13] R. Jain, Y. Shi, Z. Wei, and S. Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. *IEEE Transactions on Information Theory*, 59(8):5171–5178, 2013.

[KP16] I. Klep and J. Povh. Constrained trace-optimization of polynomials in freely noncommuting variables. *Journal of Global Optimization*, 64(2):325–348, 2016.

[KS08] I. Klep and M. Schweighofer. Connes' embedding conjecture and sums of hermitian squares. *Advances in Mathematics*, 217(4):1816–1837, 2008.

[Las01] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[Las09] J. B. Lasserre. *Moments, Positive Polynomials and Their Applications*. Imperial College Press, 2009.

[Las14] J. B. Lasserre. New approximations for the cone of copositive matrices and its dual. *Mathematical Programming*, 144(1-2):265–276, 2014.

[Lau09] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In M. Putinar and S. Sullivant, editors, *Emerging Applications of Algebraic Geometry*, pages 157–270. Springer, 2009.

[LP15] M. Laurent and T. Piovesan. Conic approach to quantum graph parameters using linear optimization over the completely positive semidefinite cone. *SIAM Journal on Optimization*, 25(4):2461–2493, 2015.

[LRS15] J. R. Lee, P. Raghavendra, and 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, pages 567–576, 2015.

[LWdW17] T. Lee, Z. Wei, and R. de Wolf. Some upper and lower bounds on psd-rank. *Mathematical Programming*, 162(1):495–521, 2017.

[Mar91] R. K. Martin. Using separation algorithms to generate mixed integer model reformulations. *Operations Research Letters*, 10(3):119–128, 1991.

[MGS81] L. Lovász M. Grötschel and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

[MR14] L. Mančinska and D. E. Roberson. Note on the correspondence between quantum correlations and the completely positive semidefinite cone. *Available at `quantuminfo.quantumlah.org/memberpages/laura/corr.pdf`*, 2014.

[MSvS03] D. Mond, J. Smith, and 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):2821–2845, 2003.

[Nie14] J. Nie. The $\mathcal{A}$-truncated K-moment problem. *Foundations of Computational Mathematics*, 14(6):1243–1276, 2014.

[Nie16] J. Nie. Symmetric tensor nuclear norms. *arXiv:1605.08823*, 2016.

[NPA12] M. Navascués, S. Pironio, and A. Acín. SDP relaxations for non-commutative polynomial optimization. In M. F. Anjos and J. B. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, pages 601–634. Springer, 2012.

[Par00] P. A. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, Caltech, 2000.

[PNA10]   S. Pironio, M. Navascués, and A. Acín. Convergent relaxations of polynomial optimization problems with noncommuting variables. *SIAM Journal on Optimization*, 20(5):2157–2180, 2010.

[PSVW16]   A. Prakash, J. Sikora, A. Varvitsiotis, and Z. Wei. Completely positive semidefinite rank. *arXiv:1604.07199*, 2016.

[Put93]   M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42:969–984, 1993.

[PV17]   A. Prakash and A. Varvitsiotis. Matrix factorizations of correlation matrices and applications. *arXiv:1702.06305*, 2017.

[Ren92]   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):255 – 299, 1992.

[Rot14]   T. Rothvoss. The matching polytope has exponential extension complexity. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing*, STOC'14, pages 263–272, 2014.

[Rud87]   W. Rudin. *Real and complex analysis.* Mathematics series. McGraw-Hill, 1987.

[Shi16]   Y. Shitov. The complexity of positive semidefinite matrix factorization. *arXiv:1606.0906*, 2016.

[Slo17]   W. Slofstra. The set of quantum correlations is not closed. *arXiv:1703.08618*, 2017.

[SMBB+15] N. Shaked-Monderer, A. Berman, I. M. Bomze, F. Jarre, and W. Schachinger. New results on the cp-rank and related properties of co(mpletely)positive matrices. *Linear and Multilinear Algebra*, 63(2):384–396, 2015.

[SMBJS13] N. Shaked-Monderer, I. M. Bomze, F. Jarre, and W. Schachinger. On the cp-rank and minimal cp factorizations of a completely positive matrix. *SIAM Journal on Matrix Analysis and Applications*, 34(2):355–368, 2013.

[SV17]   J. Sikora and A. Varvitsiotis. Linear conic formulations for two-party correlations and values of nonlocal games. *Mathematical Programming*, 162(1):431–463, 2017.

[TS15]   G. Tang and P. Shah. Guaranteed tensor decomposition: A moment approach. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 1491–1500, 2015.

[Vav09]   S. A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, 20(3):1364–1377, 2009.

[VGG17]   A. Vandaele, F. Gineur, and N. Gillis. Algorithms for positive semidefinite factorization. *arXiv:1707.07953v1*, 2017.

[Wed64]   J. H. M. Wedderburn. *Lectures on Matrices.* Dover Publications Inc., 1964.

[Yan91]   M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441 – 466, 1991.