

OPTIMIZATION OVER TRACE POLYNOMIALS

IGOR KLEP, VICTOR MAGRON, AND JURIJ VOLČIČ

ABSTRACT. Motivated by recent progress in quantum information theory, this article aims at optimizing trace polynomials, i.e., polynomials in noncommuting variables and traces of their products. A novel Positivstellensatz certifying positivity of trace polynomials subject to trace constraints is presented, and a hierarchy of semidefinite relaxations converging monotonically to the optimum of a trace polynomial subject to tracial constraints is provided. This hierarchy can be seen as a tracial analog of the Pironio, Navascués and Acín scheme [New J. Phys., 2008] for optimization of noncommutative polynomials. The Gelfand-Naimark-Segal (GNS) construction is applied to extract optimizers of the trace optimization problem if flatness and extremality conditions are satisfied. These conditions are sufficient to obtain finite convergence of our hierarchy. The results obtained are applied to violations of polynomial Bell inequalities in quantum information theory. The main techniques used in this paper are inspired by real algebraic geometry, operator theory, and noncommutative algebra.

1. INTRODUCTION

The goal of this article is to solve the class of polynomial optimization problems with noncommuting variables (e.g., polynomials in matrices) involving the trace. Applications of interest arise from quantum theory and quantum information science [NPA08, PKRR⁺19, Hub21] as well as control theory [SIG98, dOHMP09]. Further motivation relates to the generalized Lax conjecture [Lax58], where the goal is to obtain computer-assisted proofs based on noncommutative sums of squares in Clifford algebras [NT14]. The verification of noncommutative polynomial trace inequalities has also been motivated by a conjecture formulated by Bessis, Moussa and Villani (BMV) in 1975 [BMV75], which has been recently proved by Stahl [Sta13] (see also the Lieb and Seiringer reformulation [LS04]). Further efforts focused on applications arising from bipartite quantum correlations [GdLL18], and matrix factorization ranks in [GdLL19]. In a related analytic direction, there has been recent progress on multivariate generalizations of the Golden-Thompson inequality and the Araki-Lieb-Thirring inequality [SBT17, HKT17].

2010 *Mathematics Subject Classification.* 46N50; 90C22; 47N10; 13J30; 81-08.

Key words and phrases. Noncommutative polynomial; (pure) trace polynomial; semialgebraic set; semidefinite programming; trace optimization; Positivstellensatz; von Neumann algebra.

IK was supported by the Slovenian Research Agency grants J1-2453, J1-8132 and P1-0222.

VM was supported by the FMJH Program PGMO (EPICS project) and EDF, Thales, Orange et Criteo, as well as from the Tremplin ERC Stg Grant ANR-18-ERC2-0004-01 (T-COPS project).

JV was supported by the NSF grant DMS-1954709.

There is a plethora of prior research in quantum information theory involving reformulating problems as optimization of noncommutative polynomials. One famous application is to characterize the set of quantum correlations. Bell inequalities [Bel64] provide a method to investigate entanglement, which allows two or more parties to be correlated in a non-classical way, and is often studied through the set of bipartite quantum correlations. Such correlations consist of the conditional probabilities that two physically separated parties can generate by performing measurements on a shared entangled state. These conditional probabilities satisfy some inequalities classically, but violate them in the quantum realm [CHSH69].

Classically, *polynomial optimization* aims at minimizing a polynomial over a set defined by a finite conjunction of polynomial inequalities, i.e., a basic closed *semialgebraic set*. Solving this optimization problem is NP-hard in general [Lau09]. *Lasserre’s hierarchy* [Las01] is a nowadays well established methodology to handle polynomial optimization in a practical way. This framework consists of approximating the solution of the initial problem by considering a hierarchy of convex relaxations. Each step of the hierarchy boils down to computing the optimal value of a *semidefinite program* [AL12], that is, the optimum of a linear function under linear matrix inequality constraints. As a consequence of Putinar’s Positivstellensatz [Put93], if the quadratic module generated by the polynomials describing the semialgebraic set is archimedean, the hierarchy of semidefinite bounds converges from below to the minimum of the polynomial over this semialgebraic set.

In the free noncommutative context (i.e., without traces), a polynomial is positive semidefinite if and only if it can be written as a *sum of hermitian squares* (SOHS) [Hel02, McC01]. One can rely on such SOHS decompositions to perform eigenvalue optimization of noncommutative polynomials over noncommutative semialgebraic sets, i.e., under noncommutative polynomial inequality constraints. The noncommutative analogues of Lasserre’s hierarchy [HM04, NPA08, PNA10, CKP12, BCKP13] allow one to approximate as closely as desired the optimal value of such eigenvalue minimization problems. In [NPA08], Navascués, Pironio and Acín provide a way to compute bounds on the maximal violation levels of Bell inequalities: they first reformulate the initial problem as an eigenvalue optimization one and then approximate its solution with a converging hierarchy of semidefinite programs, based on the noncommutative version of Putinar’s Positivstellensatz due to Helton and McCullough [HM04]. This is the so-called Navascués-Pironio-Acín (NPA for short) hierarchy and can be viewed as the “eigenvalue” version of Lasserre’s hierarchy. This leads to a hierarchy of upper bounds on the maximum violation level of Bell inequalities (see also [DLTW08, PV09]). Further extensions [PNA10, CKP12, BCKP13] have been provided to optimize the trace of a given polynomial under positivity constraints. *NCSOStools* [CKP11, BKP16] can compute lower bounds on minimal eigenvalues or traces of noncommutative polynomial objective functions over noncommutative semialgebraic sets.

This work greatly extends these frameworks to the case of optimization problems involving *trace polynomials*, i.e., linear combinations of products of matrices and matrix traces. A very simple example of such polynomial is $\text{tr}(A_1) \cdot \text{tr}(A_2^2) + (\text{tr}(A_1 A_2))^2$, where A_1 and A_2 are noncommutative variables, e.g., A_1 and A_2 can

be both quantum physics operators. One important underlying motivation is that trace polynomials are involved in several problems arising from quantum information theory. For instance, [FN14] presents a framework to obtain the limit output states for a large class of input states having specific sets of parameters. To obtain these limits, one needs to compute bounds for generalized traces of tensors. One way to model such generalized traces is to consider a reformulation as an optimization problem involving trace polynomials. In this problem, trace polynomials arise as cost functions but they can also appear in the constraints. Convex relaxations of trace polynomial problems can be obtained as in the NPA hierarchy: one can associate a new variable to each word trace (e.g., $\text{tr}(A_1)$, $\text{tr}(A_1 A_2)$ and $\text{tr}(A_2^2)$ in the above example), then incorporate the initial constraints into the semidefinite matrix defined in the NPA hierarchy. Moreover the noncommuting operators, denoted by A_i, B_j, C_k in [PKRR⁺19], fulfill causal constraints, which leads to equality constraints such as

$$\text{tr}(A_{i_1} A_{i_2} \cdots A_{i_m} C_{k_1} C_{k_2} \cdots C_{k_m}) - \text{tr}(A_{i_1} A_{i_2} \cdots A_{i_m}) \text{tr}(C_{k_1} C_{k_2} \cdots C_{k_m}) = 0.$$

This results in a so-called *scalar extension* of the NPA hierarchy, which allows the authors to successfully identify correlations not attainable in the entanglement-swapping scenario. However, [PKRR⁺19] does not provide a proof of convergence for this hierarchy. In [Hub21], the author focuses on the multilinear case and obtains a characterization of all multilinear equivariant trace polynomials which are positive on the positive cone. In contrast with [Hub21], we consider optimization of nonlinear trace polynomials over trace polynomial inequalities after assuming that the quadratic module generated by the polynomials involved in the set of constraints is archimedean. In a closely related work in real algebraic geometry [KŠV18], the first and third author derive several Positivstellensätze for trace polynomials positive on semialgebraic sets of *fixed size* matrices. In particular, [KŠV18] establishes a Putinar-type Positivstellensatz stating that any positive polynomial admits a weighted SOHS decomposition without denominators. In the dimension-free setting, finite von Neumann algebras and their tracial states provide a natural framework for studying tracial polynomial inequalities. This paper characterizes trace polynomials which are positive on tracial semialgebraic sets, where the initial polynomials and constraints involve freely noncommutative variables and traces, and the evaluations are performed on von Neumann algebras.

Contributions. A *trace polynomial* is a polynomial in symmetric noncommutative variables x_1, \dots, x_n and traces of their products. Thus naturally each trace polynomial has an adjoint. A *pure trace polynomial* is a trace polynomial that is made only of traces, i.e., has no free variables x_j . For instance, the trace of a trace polynomial is a pure trace polynomial, e.g.

$$\begin{aligned} f &= x_1 x_2 x_1^2 - \text{tr}(x_2) \text{tr}(x_1 x_2) \text{tr}(x_1^2 x_2) x_2 x_1, \\ \text{tr}(f) &= \text{tr}(x_1^3 x_2) - \text{tr}(x_2) \text{tr}(x_1 x_2)^2 \text{tr}(x_1^2 x_2), \\ f^* &= x_1^2 x_2 x_1 - \text{tr}(x_2) \text{tr}(x_1 x_2) \text{tr}(x_1^2 x_2) x_1 x_2. \end{aligned}$$

Given a set of symmetric trace polynomials S , let \mathcal{D}_S be the set of all tuples (X_1, \dots, X_n) , where X_j are operators from a finite von Neumann algebra with a given tracial state, that satisfy $s(X_1, \dots, X_n) \succeq 0$ for all $s \in S$.

In Section 3 we state and prove a pure trace variant of the Helton-McCullough [HM04] noncommutative version of Putinar's Positivstellensatz [Put93]: we obtain a representation of pure trace polynomials positive on a set described by pure trace polynomial inequalities, using weighted sums of squares. This first (noncyclic) Positivstellensatz is valid under the classical assumption that the quadratic module generated by the polynomials involved in the set of constraints is archimedean (Corollary 3.6). Our proof relies on the classical Kadison–Dubois representation theorem (see e.g. [Mar08]) and the Gelfand-Naimark-Segal (GNS) construction.

Then, we derive in Section 4 a novel, cyclic Positivstellensatz for the more general case of trace polynomials which are positive on tracial semialgebraic sets. A subset of symmetric trace polynomials \mathcal{M}^{cyc} is a *cyclic quadratic module* if $1 \in \mathcal{M}^{\text{cyc}}$, $\mathcal{M}^{\text{cyc}} + \mathcal{M}^{\text{cyc}} \subseteq \mathcal{M}^{\text{cyc}}$, $\text{tr}(\mathcal{M}^{\text{cyc}}) \subseteq \mathcal{M}^{\text{cyc}}$ and $h\mathcal{M}^{\text{cyc}}h^* \subseteq \mathcal{M}^{\text{cyc}}$ for every trace polynomial h . In analogy with the commutative setting we say that \mathcal{M}^{cyc} is *archimedean* if $N - x_1^2 - \dots - x_n^2 \in \mathcal{M}^{\text{cyc}}$ for some $N > 0$. Observe that each \mathcal{M}^{cyc} contains all sums of elements of the form $\text{tr}(h_1 h_1^*) \cdots \text{tr}(h_\ell h_\ell^*) h_0 h_0^*$ for $h_i \in \mathbb{T}$. Lemma 4.1 below describes the smallest cyclic quadratic module $\mathcal{M}^{\text{cyc}}(S)$ containing a given set of generators $S \subseteq \mathbb{T}$.

Theorem A (Corollary 4.8). *Let \mathcal{M}^{cyc} be an archimedean cyclic quadratic module, and let a be a symmetric trace polynomial. The following are equivalent:*

- (i) $a \succeq 0$ on $\mathcal{D}_{\mathcal{M}^{\text{cyc}}}$;
- (ii) for every $\varepsilon > 0$ there exist univariate sums of squares $s_1, s_2 \in \mathbb{R}[t]$, depending on ε , such that

$$a = s_1(a) - s_2(a), \quad \varepsilon - \text{tr}(s_2(a)) \in \mathcal{M}^{\text{cyc}}.$$

In Section 5, we rely on this Positivstellensatz to design a converging hierarchy of semidefinite relaxations to approximate from below the minimum of a pure trace polynomial under pure trace polynomial inequality constraints. An extension of this hierarchy to the more general case of trace polynomial constraints is presented in Section 5.3.

Theorem B (Corollary 5.7). *Let S be a set of symmetric trace polynomials, and a a pure trace polynomial. The Positivstellensatz-induced hierarchy of semidefinite programs produces a convergent increasing sequence with limit $\inf_{\mathcal{D}_S} a$.*

Along the way, we present in Section 5.2 a tracial variant of the finite-dimensional GNS construction under flatness and extremal assumptions. We use it to obtain finite convergence of our hierarchy as well as exactness of the relaxed solution, and design an algorithm to extract minimizers. Finally, in Section 6 we give a simple example demonstrating our theoretical results, and present an application of our techniques to quantum information theory: we use tracial optimization to find upper bounds on violations of polynomial Bell inequalities.

2. NOTATION AND BASIC DEFINITIONS

We begin by introducing basic notions about noncommutative polynomials, trace polynomials, and semialgebraic sets that will be used throughout the paper.

2.1. Noncommutative polynomials and trace polynomials. Let us denote by \mathbb{M}_k (resp. \mathbb{S}_k) the space of all real (resp. symmetric) matrices of order k . The normalized trace of a matrix $A \in \mathbb{M}_k$ is given by $\text{tr } A = \frac{1}{k} \sum_{i=1}^k A_{i,i}$. For a fixed $n \in \mathbb{N}$, we consider a finite alphabet x_1, \dots, x_n and generate all possible words of finite length in these letters. The empty word is denoted by 1. The resulting set of words is the *free monoid* $\langle \underline{x} \rangle$, with $\underline{x} = (x_1, \dots, x_n)$. We denote by $\mathbb{R}\langle \underline{x} \rangle$ the set of real polynomials in noncommutative variables, abbreviated as *nc polynomials*. The algebra $\mathbb{R}\langle \underline{x} \rangle$ is equipped with the involution \star that fixes $\mathbb{R} \cup \{x_1, \dots, x_n\}$ point-wise and reverses words, so that $\mathbb{R}\langle \underline{x} \rangle$ is the \star -algebra freely generated by n symmetric letters x_1, \dots, x_n .

We now introduce some algebraic terminology to deal with the trace, following [Pro76] (see also [KŠ17, KŠV18]). Two words $u, v \in \langle \underline{x} \rangle$ are called \star -cyclically equivalent ($u \stackrel{\text{cyc}^\star}{\sim} v$) if v or v^\star can be obtained from u by cyclically rotating the letters in u . For example, all words of length 3 are \star -cyclically equivalent, but $x_1x_2x_3x_4 \stackrel{\text{cyc}^\star}{\not\sim} x_2x_1x_3x_4$. We denote by \mathbb{T} the commutative polynomial algebra in infinitely many variables $\text{tr}(w)$ with $w \in \langle \underline{x} \rangle$, up to \star -cyclic equivalence, that is, $\mathbb{T} := \mathbb{R}[\text{tr}(w), w \in \langle \underline{x} \rangle / \text{cyc}^\star]$. We also let $\mathbb{T} := \mathbb{T}\langle \underline{x} \rangle$ be the free \mathbb{T} -algebra on \underline{x} . Elements of \mathbb{T} are called *pure trace polynomials*, and elements of \mathbb{T} are *trace polynomials*. For example, $t = \text{tr}(x_1^2) - \text{tr}(x_1)^2 \in \mathbb{T}$ and $x_1^2 - \text{tr}(x_1)x_1 - 2t \in \mathbb{T} = \mathbb{T}\langle x_1 \rangle$. The involution on \mathbb{T} , denoted also by \star , fixes $\{x_1, \dots, x_n\} \cup \mathbb{T}$ point-wise, and reverses words from $\langle \underline{x} \rangle$. The set of all *symmetric elements* of \mathbb{T} is defined as $\text{Sym } \mathbb{T} := \{f \in \mathbb{T} : f = f^\star\}$. A linear functional $L : \mathbb{T} \rightarrow \mathbb{R}$ is said to be *tracial* if $L(\text{tr}(f)) = L(f)$ for all $f \in \mathbb{T}$. We also consider the universal trace map τ defined by

$$\begin{aligned} \tau : \mathbb{T} &\rightarrow \mathbb{T}, \\ f &\mapsto \text{tr}(f). \end{aligned}$$

A linear functional $L : \mathbb{T} \rightarrow \mathbb{R}$ is tracial if and only if $L \circ \tau = L$. Such an L is determined by $L|_{\mathbb{T}} : \mathbb{T} \rightarrow \mathbb{R}$ being an (arbitrary) linear functional. The functional L is called *unital* if $L(1) = 1$ and is called *symmetric* if $L(f^\star) = L(f)$, for all f belonging to the domain of L .

2.2. Tracial semialgebraic sets and von Neumann algebras. Given $S \subseteq \text{Sym } \mathbb{T}$, the *matricial tracial semialgebraic set* \mathcal{D}_S associated to S is defined as follows:

$$(2.1) \quad \mathcal{D}_S := \bigcup_{k \in \mathbb{N}} \{ \underline{A} = (A_1, \dots, A_n) \in \mathbb{S}_k^n : s(\underline{A}) \succeq 0 \text{ for all } s \in S \}.$$

While (2.1) looks like a natural candidate for testing positivity of tracial polynomials, the failure of Connes' embedding conjecture [JNV⁺20] hinders the existence of a reasonable Positivstellensatz for (2.1) by [KS08]. Instead of just matrices of

all finite sizes, one is thus led to include bounded operators, similarly as in the trace-free setting [HM04]. Since we deal with tracial constraints, the considered bounded operators need to admit traces. The natural framework is therefore given by tracial von Neumann algebras, which we discuss next.

A real von Neumann algebra \mathcal{F} [ARU97] is a unital, weakly closed, real, self-adjoint subalgebra of the (real) algebra of bounded linear operators on a complex Hilbert space, with the property $\mathcal{F} \cap i\mathcal{F} = \{0\}$. We restrict ourselves to separable Hilbert spaces, implying that all von Neumann algebras have separable preduals. Much of the structure theory of real von Neumann algebras can be transferred from complex von Neumann algebras [Tak02, Chapter 5]. Namely, the complexification of a real von Neumann algebra yields a complex von Neumann algebra with an involutory $*$ -antiautomorphism; conversely, the fixed set of an involutory $*$ -antiautomorphism on a complex von Neumann algebra is a real von Neumann algebra. A real von Neumann algebra is *finite* if in its complexification, every isometry is a unitary. By [Tak02, Theorem 2.4], a von Neumann algebra is finite if and only if it admits sufficiently many normal tracial states, which will play an important role in this article.

A (real) von Neumann algebra is a factor if its center consists of only the (real) scalar operators. By [Tak02, Theorem 2.6], a factor is finite if and only if it admits a faithful normal tracial state; in this case, such a state is unique, and is called the *trace* of the factor. Finally, a II_1 -factor is an infinite-dimensional finite factor (other finite factors are of type I_n , which are $n \times n$ complex matrices in the complex setting, and $n \times n$ real matrices or $\frac{n}{2} \times \frac{n}{2}$ quaternion matrices in the real setting). In this article we consider positivity on operator semialgebraic sets. These are defined as follows (cf. [BKP16, Definition 1.59]):

Definition 2.1. *A tracial pair (\mathcal{F}, τ) consists of a real finite von Neumann algebra \mathcal{F} and a faithful normal tracial state τ on \mathcal{F} [Tak02, Chapter 5].*

Given $S \subseteq \text{Sym } \mathbb{T}$ let $\mathcal{D}_S^{\mathcal{F}, \tau}$ be the set of all self-adjoint tuples $\underline{X} = (X_1, \dots, X_n) \in \mathcal{F}^n$ making $s(\underline{X})$ a positive semidefinite operator for every $s \in S$; here tr is evaluated as τ . The von Neumann semialgebraic set \mathcal{D}_S^{vN} generated by S is defined as

$$\mathcal{D}_S^{vN} := \bigcup_{(\mathcal{F}, \tau)} \mathcal{D}_S^{\mathcal{F}, \tau},$$

where the union is over all tracial pairs (\mathcal{F}, τ) . Analogously, we define

$$\mathcal{D}_S^{\text{II}_1} := \bigcup_{\mathcal{F}} \mathcal{D}_S^{\mathcal{F}},$$

where the union is over all II_1 -factors (which come equipped with unique traces).

Note that finiteness of S is not needed at this stage. Unlike in the free case [HKM11], these tracial semialgebraic sets are closed neither under direct sums nor reducing subspace compressions; for example, if $s = \text{tr}(x_1) \text{tr}(x_2)$, then

$$\begin{aligned} s(3, 1) > 0 \quad \text{and} \quad s(-1, -2) > 0, \quad \text{but} \quad s(3 \oplus -1, 1 \oplus -2) < 0; \\ s(-2 \oplus 1, 1 \oplus -2) > 0, \quad \text{but} \quad s(-2, 1) < 0 \quad \text{and} \quad s(1, -2) < 0. \end{aligned}$$

To sidestep this technical problem we make use of the following well-known fact that is all but stated in [Dyk94, Theorem 2.5].

Proposition 2.2. *Every tracial pair embeds into a II_1 -factor.*

Proof. Let (\mathcal{F}, τ) be a tracial pair and $(\tilde{\mathcal{F}}, \tilde{\tau})$ its complexification. If $L(\mathbb{Z})$ is the complex von Neumann group algebra of \mathbb{Z} [Tak02, Definition 7.4], the free product construction (see e.g. [Dyk94, Section 1]) along $\tilde{\tau}$ and the trace on $L(\mathbb{Z})$ yields the von Neumann algebra $\tilde{\mathcal{F}} * L(\mathbb{Z})$ with a normal faithful tracial state whose restriction to $\tilde{\mathcal{F}}$ equals $\tilde{\tau}$. Also, both $L(\mathbb{Z})$ and $\tilde{\mathcal{F}}$ admit natural involutory \star -antiautomorphisms, which induce an involutory \star -antiautomorphism $\tilde{\mathcal{F}} * L(\mathbb{Z})$. Its fixed set is a real von Neumann algebra containing (\mathcal{F}, τ) , and it is a (real) II_1 -factor if $\tilde{\mathcal{F}} * L(\mathbb{Z})$ is a (complex) II_1 -factor. If $\dim \tilde{\mathcal{F}} \leq 2$, then $\tilde{\mathcal{F}} \in \{\mathbb{C}, \mathbb{C} \oplus \mathbb{C}\}$ and $\tilde{\mathcal{F}} * L(\mathbb{Z}) \in \{L(\mathbb{Z}), L(\mathbb{Z}_2 * \mathbb{Z})\}$ is a II_1 -factor. If $\dim \tilde{\mathcal{F}} \geq 3$, then $\tilde{\mathcal{F}} * L(\mathbb{Z})$ is a II_1 -factor by [Dyk94, Theorem 2.5] (and the proof of [Dyk94, Lemma 2.9]). \square

Remark 2.3. *A reader might rightfully wonder why the setup is restricted to reals instead of complexes. Since every complex von Neumann algebra is a real von Neumann algebra and the real part of a complex tracial state is a real tracial state, all the conclusions in this paper also hold for evaluations in complex von Neumann algebras. Likewise one could consider complex trace polynomials, but the corresponding formalism for trace symbols would be more intricate (namely, trace symbols would not be fixed under the involution, so they would need to be split in real and imaginary part, and relations connecting both with respect to \star would need to be imposed). However, with the view towards optimization and implementation using the standard semidefinite programming solvers it is more convenient to derive results within the real framework.*

3. NON-CYCLIC POSITIVSTELLENSATZ FOR PURE TRACE POLYNOMIALS

In this section we provide our first Positivstellensatz, Theorem 3.5, for pure trace polynomials based on quadratic modules from real algebraic geometry [Mar08]. A subset $\mathcal{M} \subseteq \mathbb{T}$ is called a quadratic module if $1 \in \mathcal{M}$, $\mathcal{M} + \mathcal{M} \subseteq \mathcal{M}$ and $a^2\mathcal{M}$ for all $a \in \mathbb{T}$. Given an archimedean quadratic module $\mathcal{M} \subseteq \mathbb{T}$ (in the usual commutative sense, meaning that for each $f \in \mathbb{T}$ there is $m > 0$ such that $m \pm f \in \mathcal{M}$), we consider the real points of the real spectrum $\text{Sper}_{\mathcal{M}} \mathbb{T}$, namely the set $\chi_{\mathcal{M}}$ defined by

$$(3.1) \quad \chi_{\mathcal{M}} := \{\varphi : \mathbb{T} \rightarrow \mathbb{R} \mid \varphi \text{ homomorphism, } \varphi(\mathcal{M}) \subseteq \mathbb{R}_{\geq 0}, \varphi(1) = 1\}.$$

The next proposition is the well-known Kadison-Dubois representation theorem, see e.g. [Mar08, Theorem 5.4.4].

Proposition 3.1. *Let $\mathcal{M} \subseteq \mathbb{T}$ be an archimedean quadratic module. Then, for all $a \in \mathbb{T}$, one has*

$$\forall \varphi \in \chi_{\mathcal{M}} \quad \varphi(a) \geq 0 \quad \Leftrightarrow \quad \forall \varepsilon > 0 \quad a + \varepsilon \in \mathcal{M}.$$

A homomorphism $\varphi : \mathbb{T} \rightarrow \mathbb{R}$ is determined by the ‘‘tracial moments’’ $\varphi(\text{tr}(w))$ for $w \in \langle \underline{x} \rangle$. In this sense, the following variant of [Had01, Theorem 1.3] is a solution

of the tracial moment problem. In the given formulation, it is the dimension-free analog of the extension theorem [KŠV18, Theorem 4.8].

Proposition 3.2. *Let $\varphi : \mathbb{T} \rightarrow \mathbb{R}$ be a homomorphism. Then there are a tracial pair (\mathcal{F}, τ) and $\underline{X} = \underline{X}^* \in \mathcal{F}^n$ such that $\varphi(a) = a(\underline{X})$ for all $a \in \mathbb{T}$ if and only if the following holds:*

- (a) $\varphi(\text{tr}(pp^*)) \geq 0$ for all $p \in \mathbb{R}\langle \underline{x} \rangle$;
- (b) $\liminf_{k \rightarrow \infty} \sqrt[2k]{\varphi(\text{tr}(x_j^{2k}))} < \infty$ for $j = 1, \dots, n$.

Proof. (\Rightarrow) This is trivial since $\tau(AA^*) \geq 0$ and $|\tau(A^{2k})| \leq \|A\|^{2k}$ for every $A \in \mathcal{F}$ and $k \in \mathbb{N}$.

(\Leftarrow) Denote

$$\alpha = \max \left\{ 1, \max_j \liminf_{k \rightarrow \infty} \sqrt[2k]{\varphi(\text{tr}(x_j^{2k}))} \right\}.$$

Let $\phi : \mathbb{R}\langle \underline{x} \rangle \rightarrow \mathbb{R}$ be a linear functional defined by

$$\phi(w) := \frac{\varphi(\text{tr}(w))}{\alpha^{|w|}}$$

for $w \in \mathbb{R}\langle \underline{x} \rangle$. Then ϕ is a symmetric tracial functional on $\mathbb{R}\langle \underline{x} \rangle$, $\phi(pp^*) \geq 0$ for every $p \in \mathbb{R}\langle \underline{x} \rangle$ and $\max_j \liminf_{k \rightarrow \infty} \phi(x_j^{2k}) < \infty$. By [Had01, Theorem 1.3] (or rather its real version) there is a tracial pair (\mathcal{F}, τ) and a tuple of self-adjoint contractions $\underline{Y} \in \mathcal{F}^n$ such that $\phi(p) = \tau(p(\underline{Y}))$ for all $p \in \mathbb{R}\langle \underline{x} \rangle$. Then $\underline{X} = \alpha \underline{Y}$ satisfies $\varphi(\text{tr}(p)) = \tau(p(\underline{X}))$ for $p \in \mathbb{R}\langle \underline{x} \rangle$ and thus $\varphi(a) = a(\underline{X})$ for $a \in \mathbb{T}$. \square

Definition 3.3. *Given $S \subseteq T$ and $N > 0$ let*

$$(3.2) \quad S(N) := S \cup \{ \text{tr}(pp^*) \mid p \in \mathbb{R}\langle \underline{x} \rangle \} \cup \{ N^k - \text{tr}(x_j^{2k}) \mid 1 \leq j \leq n, k \in \mathbb{N} \} \subseteq T.$$

For $S \subseteq \text{Sym } \mathbb{T}$ let

$$(3.3) \quad S[N] := S \cup \{ N - x_j^2 \mid j = 1, \dots, n \} \subseteq \mathbb{T}.$$

Lemma 3.4. *The quadratic module $\mathcal{M}(S(N)) \subseteq \mathbb{T}$ is archimedean for every S, N .*

Proof. We need to show that for every $w \in \mathbb{R}\langle \underline{x} \rangle$ there exists $m > 0$ such that

$$(3.4) \quad m \pm \text{tr}(w) \in \mathcal{M}(S(N)).$$

Write $w = x_{i_1}^{k_1} \cdots x_{i_\ell}^{k_\ell}$ for $i_1 \neq i_2 \neq \cdots \neq i_\ell$; we prove (3.4) by induction on ℓ . If $\ell = 1$, (3.4) holds because

$$N^k + 1 \pm 2 \text{tr}(x_j^k) = (N^k - \text{tr}(x_j^{2k})) + \text{tr}((x_j^k \pm 1)^2).$$

For $\ell > 1$ denote $\lambda = \lfloor \frac{\ell}{2} \rfloor$, and let $w_1 = x_{i_1}^{k_1} \cdots x_{i_\lambda}^{k_\lambda}$ and $w_2 = x_{i_{\lambda+1}}^{k_{\lambda+1}} \cdots x_{i_\ell}^{k_\ell}$. Then

$$\text{tr}(w_1 w_1^*) = \text{tr} \left(x_{i_1}^{2k_1} x_{i_2}^{2k_2} \cdots x_{i_{\lambda-1}}^{2k_{\lambda-1}} x_{i_\lambda}^{2k_\lambda} x_{i_{\lambda-1}}^{2k_{\lambda-1}} \cdots x_{i_2}^{2k_2} \right)$$

and similarly for $\text{tr}(w_2 w_2^*)$; note that $2(\lambda - 1), 2(\ell - \lambda - 1) < \ell$. Hence by the induction hypothesis there exist $m_1, m_2 > 0$ such that $m_i - \text{tr}(w_i w_i^*) \in \mathcal{M}(S(N))$. Then $(m_1 + m_2) \pm 2 \text{tr}(w) = (m_1 - \text{tr}(w_1 w_1^*)) + (m_2 - \text{tr}(w_2 w_2^*)) + \text{tr}((w_1 \pm w_2^*)(w_1^* \pm w_2))$ lies in $\mathcal{M}(S(N))$. \square

Recall that the Helton-McCullough archimedean Positivstellensatz [HM04] states that any polynomial in noncommutative variables positive on a basic semialgebraic set belongs to the quadratic module generated by the polynomials describing this set, under the assumption that this module is archimedean. We are now ready to prove our first theorem, the purely tracial analog of this noncommutative Helton-McCullough Positivstellensatz.

Theorem 3.5. *Let $S \subseteq \mathbb{T}$ and $N > 0$ be given. Then for $a \in \mathbb{T}$ the following are equivalent:*

- (i) $a(\underline{X}) \geq 0$ for all $\underline{X} \in \mathcal{D}_{S[N]}^{vN}$;
- (ii) $a(\underline{X}) \geq 0$ for all $\underline{X} \in \mathcal{D}_{S[N]}^{\text{II}_1}$;
- (iii) $a + \varepsilon \in \mathcal{M}(S(N))$ for all $\varepsilon > 0$.

Proof. (i) \Leftrightarrow (ii) holds by Proposition 2.2. (iii) \Rightarrow (i) If $\underline{X} \in \mathcal{D}_{S[N]}^{\mathcal{F},\tau}$, then

$$s(\underline{X}) \geq 0, \quad \tau(p(X)p(X)^*) \geq 0, \quad \tau(X_j^{2k}) \leq N^k$$

for all $s \in S$, $p \in \mathbb{R}\langle x \rangle$, $1 \leq j \leq n$ and $k \in \mathbb{N}$, so $a(\underline{X}) \geq 0$.

(i) \Rightarrow (iii) Suppose $a + \varepsilon \notin \mathcal{M}(S(N))$ for some $\varepsilon > 0$. By Proposition 3.1, there exists a unital homomorphism $\varphi : \mathbb{T} \rightarrow \mathbb{R}$ with $\varphi(\mathcal{M}(S(N))) \subseteq \mathbb{R}_{\geq 0}$ and $\varphi(a) < 0$. Hence

$$\varphi(\text{tr}(pp^*)) \geq 0, \quad \varphi(\text{tr}(x_j^{2k})) \leq N^k$$

for all $p \in \mathbb{R}\langle x \rangle$, $1 \leq j \leq n$ and $k \in \mathbb{N}$. Hence by Proposition 3.2 there exist a tracial pair (\mathcal{F}, τ) and $\underline{X} \in \mathcal{F}^n$ such that $\varphi(b) = b(\underline{X})$ for all $b \in \mathbb{T}$. Moreover, the proof of Proposition 3.2 implies $\|X_j\| \leq \sqrt{N}$ for $1 \leq j \leq n$. Furthermore, $s(\underline{X}) = \varphi(s) \geq 0$ for every $s \in S$ implies $\underline{X} \in \mathcal{D}_{S[N]}^{\mathcal{F},\tau}$. Finally, $a(\underline{X}) = \varphi(a) < 0$. \square

Since $\mathcal{M}(S(N_1)) \subseteq \mathcal{M}(S(N_2))$ for $N_1 \geq N_2$, we obtain the following:

Corollary 3.6. *Let $S \subseteq \mathbb{T}$ and $a \in \mathbb{T}$. The following are equivalent:*

- (i) $a(\underline{X}) \geq 0$ for all $\underline{X} \in \mathcal{D}_S^{vN}$;
- (ii) $a(\underline{X}) \geq 0$ for all $\underline{X} \in \mathcal{D}_S^{\text{II}_1}$;
- (iii) $a + \varepsilon \in \mathcal{M}(S(N))$ for all $\varepsilon > 0$ and $N \in \mathbb{N}$.

4. CYCLIC POSITIVSTELLENSATZ FOR TRACE POLYNOMIALS

In this section we prove a Positivstellensatz for trace polynomials that is less inspired by the commutative theory than the one from Section 3 and relies more on the tracial structure of trace polynomials. First we introduce the notion of a cyclic quadratic module. A subset $\mathcal{M}^{\text{cyc}} \subseteq \text{Sym } \mathbb{T}$ is called a *cyclic quadratic module* if

$$1 \in \mathcal{M}^{\text{cyc}}, \quad \mathcal{M}^{\text{cyc}} + \mathcal{M}^{\text{cyc}} \subseteq \mathcal{M}^{\text{cyc}}, \quad a^* \mathcal{M}^{\text{cyc}} a \subseteq \mathcal{M}^{\text{cyc}} \quad \forall a \in \mathbb{T}, \quad \text{tr}(\mathcal{M}^{\text{cyc}}) \subseteq \mathcal{M}^{\text{cyc}}.$$

Given $S \subseteq \mathbb{T}$ let $\mathcal{M}^{\text{cyc}}(S)$ be the cyclic quadratic module generated by S , i.e., the smallest cyclic quadratic module in \mathbb{T} containing S . A cyclic quadratic module \mathcal{M}^{cyc} is called *archimedean* if for all $a \in \text{Sym } \mathbb{T}$ there exists $N > 0$ such that $N - a \in \mathcal{M}^{\text{cyc}}$. We start with a few preliminary results.

Lemma 4.1. *Let $S \subseteq \mathbb{T}$.*

(1) Elements of $\mathcal{M}^{\text{cyc}}(\emptyset)$ are precisely sums of

$$\text{tr}(h_1 h_1^*) \cdots \text{tr}(h_\ell h_\ell^*) h_0 h_0^*$$

for $h_i \in \mathbb{T}$.

(2) Elements of $\mathcal{M}^{\text{cyc}}(S)$ are precisely sums of

$$q_1, \quad h_1 s_1 h_1^*, \quad \text{tr}(h_2 s_2 h_2^*) q_2$$

for $h_i \in \mathbb{T}$, $q_i \in \mathcal{M}^{\text{cyc}}(\emptyset)$, $s_i \in S$.

(3) Elements of $\text{tr}(\mathcal{M}^{\text{cyc}}(S)) = \mathcal{M}^{\text{cyc}}(S) \cap T$ are precisely sums of

$$\text{tr}(h_1 h_1^*) \cdots \text{tr}(h_\ell h_\ell^*) \text{tr}(h_0 s h_0^*)$$

for $h_i \in \mathbb{T}$ and $s \in S$.

Proof. Straightforward. \square

Note that expressions such as $\text{tr}(h_1 s_1 h_1^*) \text{tr}(h_2 s_2 h_2^*)$ for $h_i \in \mathbb{T}$ and $s_i \in S$ do not belong to $\mathcal{M}^{\text{cyc}}(S)$. We emphasize that computing such product representations in our context would be very hard in practice. Indeed, even if one bounds the degrees of the h_j , computing their coefficients boils down to solving a nonlinear semidefinite program, which is impractical. Second, even in the classical commutative case quadratic modules (such as those appearing in Putinar's Positivstellensatz) are not closed under multiplication, by contrast with Schmüßgen type representations [Sch91] in which one allows multiplication of polynomials involved in the set of constraints. Admitting products of constraints in sums of squares positivity certificates increases computational cost only modestly, but from a theoretical viewpoint yields little to no advantages (in the archimedean case), which is why quadratic modules are preferred from a practical point of view.

Proposition 4.2. *A cyclic quadratic module \mathcal{M}^{cyc} is archimedean if and only if there exists $N \in \mathbb{N}$ such that $N - \sum_{i=1}^n x_j^2 \in \mathcal{M}^{\text{cyc}}$.*

Proof. (\Rightarrow) is obvious. For the converse assume $N - \sum_{i=1}^n x_j^2 \in \mathcal{M}^{\text{cyc}}$ for some $N \in \mathbb{N}$. Then the set $\mathcal{M}^{\text{cyc}} \cap \mathbb{R}\langle \underline{x} \rangle$ is an archimedean quadratic module. Thus, for all $a = a^* \in \mathbb{R}\langle \underline{x} \rangle$ there exists $N \in \mathbb{N}$ such that

$$(4.1) \quad N - a \in \mathcal{M}^{\text{cyc}} \cap \mathbb{R}\langle \underline{x} \rangle.$$

In addition, the set H of bounded elements, defined by

$$H = \{a \in \mathbb{T} \mid \exists N \in \mathbb{N} \text{ s.t. } N - a^* a \in \mathcal{M}^{\text{cyc}}\},$$

is closed under involution, addition, subtraction and multiplication, i.e., is a \star -subalgebra of \mathbb{T} [Vid59]. A symmetric element $b \in \mathbb{T}$ is in H if and only if there is some $N \in \mathbb{N}$ with $N \pm b \in \mathcal{M}^{\text{cyc}}$.

For every $a \in \langle \underline{X} \rangle$ we have

$$(4.2) \quad \text{tr}(aa^*) - \text{tr}(a)^2 = \text{tr}((a - \text{tr}(a))(a - \text{tr}(a))^*) \in \mathcal{M}^{\text{cyc}}.$$

By (4.1) and the fact that \mathcal{M}^{cyc} is cyclic, there is some $N \in \mathbb{N}$ with $N - \text{tr}(aa^*) \in \mathcal{M}^{\text{cyc}}$. Adding this to (4.2) yields $N - \text{tr}(a)^2 \in \mathcal{M}^{\text{cyc}}$. Thus, by the definition of H , $\text{tr}(a) \in \mathcal{M}^{\text{cyc}}$. The desired result now follows since H is a subalgebra of \mathbb{T} . \square

Proposition 4.3. *Let (\mathcal{F}, τ) be a tracial pair and $X = X^* \in \mathcal{F}$. The following are equivalent:*

- (i) $X \succeq 0$;
- (ii) $\tau(XY) \geq 0$ for all positive semidefinite contractions $Y \in \mathcal{F}$;
- (iii) $\tau(Xp(X)^2) \geq 0$ for all $p \in \mathbb{R}[t]$.

Proof. (i) \Rightarrow (ii) is clear and (ii) \Rightarrow (iii) holds since for $p(X) \neq 0$,

$$\tau(Xp(X)^2) = \|p(X)\|^2 \tau\left(X (\|p(X)\|^{-2} p(X)^2)\right)$$

and $\|p(X)\|^{-2} p(X)^2$ is a positive semidefinite contraction. To prove (iii) \Rightarrow (i), let $\mathcal{F}_1 \subseteq \mathcal{F}$ be the weak operator topology closure of the algebra generated by X . Then \mathcal{F}_1 is an abelian von Neumann algebra and therefore $(\mathcal{F}_1, \tau|_{\mathcal{F}_1}) \cong (L^\infty(\mathcal{X}, \mu), \int \cdot d\mu)$ for some standard measure space (\mathcal{X}, μ) by [Tak02, Theorem III.1.18]. For $f \in L^\infty(\mathcal{X}, \mu)$ we have $f \succeq 0$ if and only if $\int fg^2 d\mu \geq 0$ for all $g \in L^\infty(\mathcal{X}, \mu)$. If f is the image of X under the above isomorphism, then $\{p(f) \mid p \in \mathbb{R}[t]\}$ is dense in \mathcal{F}_1 . Hence $\int fp(f)^2 d\mu \geq 0$ for all univariate polynomials p implies $f \succeq 0$, so (iii) \Rightarrow (i) holds. \square

The following is the cyclic version of the Helton-McCullough theorem [HM04]. Note that while the constraints in Theorem 4.4 are arbitrary trace polynomials, the objective function needs to be a pure trace polynomial. A direct analog for non-pure trace objective polynomials fails, see Example 4.6 below.

Theorem 4.4. *Let $\mathcal{M}^{\text{cyc}} \subseteq \text{Sym } \mathbb{T}$ be an archimedean cyclic quadratic module and $a \in \mathbb{T}$. The following are equivalent:*

- (i) $a(\underline{X}) \geq 0$ for all $\underline{X} \in \mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{vN}$;
- (ii) $a(\underline{X}) \geq 0$ for all $\underline{X} \in \mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{\text{II}_1}$;
- (iii) $a + \varepsilon \in \mathcal{M}^{\text{cyc}}$ for all $\varepsilon > 0$.

Proof. Implications (iii) \Rightarrow (i) \Rightarrow (ii) are straightforward, so consider (ii) \Rightarrow (iii). By Proposition 4.3 we have $\mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{\text{II}_1} = \mathcal{D}_{\text{tr}(\mathcal{M}^{\text{cyc}})}^{\text{II}_1}$. Since \mathcal{M}^{cyc} is archimedean, there exists $N > 0$ such that $N - x_j^2 \in \mathcal{M}^{\text{cyc}}$ for $j = 1, \dots, n$. Consequently

$$\mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{\text{II}_1} = \mathcal{D}_{\mathcal{M}^{\text{cyc}}[N]}^{\text{II}_1} = \mathcal{D}_{\text{tr}(\mathcal{M}^{\text{cyc}})[N]}^{\text{II}_1}.$$

For every j, k we have

$$N^k - \text{tr}(x_j^{2k}) = \text{tr}\left(\left(\sum_{i=0}^{k-1} N^{k-1-i} x_j^{2i}\right)(N - x_j^2)\right) \in \mathcal{M}^{\text{cyc}}.$$

Consequently $\text{tr}(\mathcal{M}^{\text{cyc}})(N) \subseteq \mathcal{M}^{\text{cyc}}$ and thus $\mathcal{M}(\text{tr}(\mathcal{M}^{\text{cyc}})(N)) \subseteq \mathcal{M}^{\text{cyc}}$. Then $a + \varepsilon \in \mathcal{M}^{\text{cyc}}$ for all $\varepsilon > 0$ by Theorem 3.5. \square

For the reader unfamiliar with real algebraic geometry and noncommutative moment problems, we present a self-contained proof of Theorem 4.4 relying only on convex separation results and basic properties of von Neumann algebras in Appendix A.

Having reached this point, it is tempting to think that membership in \mathcal{M}^{cyc} is also enough to characterize all positive tracial polynomials, not just the pure ones.

As it turns out, this intuition is wrong: in the following lines, we provide a simple counterexample. Given a set of symmetric polynomials $S \subseteq \mathbb{R}\langle \underline{x} \rangle$ let $\mathcal{M}(S)$ denote the (free) quadratic module generated by S [BKP16, Section 1.4]. Hence $\mathcal{M}(S)$ is the smallest set that contains $S \cup \{1\}$, is closed under addition, and $f \in \mathcal{M}(S)$ implies $hfh^* \in \mathcal{M}(S)$ for every $h \in \mathbb{R}\langle \underline{x} \rangle$.

Lemma 4.5. *Let $S_1 \subseteq T$ and $S_2 \subseteq \mathbb{R}\langle \underline{x} \rangle$. If $s(0) \geq 0$ for all $s \in S_1$, then*

$$\mathcal{M}^{\text{cyc}}(S_1 \cup S_2) \cap \mathbb{R}\langle \underline{x} \rangle = \mathcal{M}(S_2).$$

Proof. Let $\pi_1 : \mathbb{T} \rightarrow \mathbb{R}$ be given by $\pi_1(a) = a(0)$, and consider the homomorphism $\pi = \pi_1 \otimes \text{id}_{\mathbb{R}\langle \underline{x} \rangle} : \mathbb{T} \rightarrow \mathbb{R}\langle \underline{x} \rangle$. Then $\pi(\mathcal{M}^{\text{cyc}}(S_1 \cup S_2)) = \mathcal{M}(S_2)$ because $\pi_1(S_1) \subseteq \mathbb{R}_{\geq 0}$ and $\pi|_{\mathbb{R}\langle \underline{x} \rangle} = \text{id}_{\mathbb{R}\langle \underline{x} \rangle}$. So the statement follows. \square

Example 4.6. *Let $n = 1$. Let \mathcal{M}^{cyc} be the archimedean cyclic quadratic module in $\text{Sym } \mathbb{T}$ generated by*

$$\{1 - x_1^2\} \cup \{\text{tr}(x_1 p^2(x_1)) \mid p \in \mathbb{R}[t]\}.$$

By Proposition 4.3, $X_1 \in \mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{\mathcal{F}, \tau}$ implies $X_1 \succeq 0$ for any tracial pair (\mathcal{F}, τ) . On the other hand, if $\varepsilon \in [0, 1)$ then $x_1 + \varepsilon \notin \mathcal{M}(\{1 - x_1^2\})$ and therefore $x_1 + \varepsilon \notin \mathcal{M}^{\text{cyc}}$ by Lemma 4.5.

We emphasize that Example 4.6 shows that Theorem 4.4 does not hold in general for non-pure objectives. To mitigate the absence of a non-pure analog of Theorem 4.4, we require the following technical lemma.

Lemma 4.7. *Let $\varepsilon > 0$ and $n = \lceil 1/\varepsilon \rceil$. If $s_2 = \frac{\varepsilon}{2}(t-1)^{2n}$ and $s_1 = s_2 + t$, then*

- (a) s_1 is positive on \mathbb{R} , and thus a sum of (two) squares in $\mathbb{R}[t]$;
- (b) $\frac{\varepsilon}{2} - s_2$ is nonnegative on $[0, 1]$, and thus an element of $\mathcal{M}(\{t, 1-t\})$.

Proof. (a) Clearly $s_1(\alpha) > 0$ for $\alpha \geq 0$. Since $\frac{ds_2}{d\alpha}(\alpha) = \varepsilon n(\alpha-1)^{2n-1} < -1$ for every $\alpha \leq 0$, we also have $s_2(\alpha) > -\alpha$ for $\alpha \leq 0$. So s_1 is positive on \mathbb{R} and thus a sum of two squares by an easy application of the fundamental theorem of algebra (see e.g. [Mar08, Proposition 1.2.1]).

(b) $\frac{\varepsilon}{2} - s_2$ is nonnegative on $[0, 1]$ because $(\alpha-1)^{2n} \leq 1$ for $\alpha \in [0, 1]$. Since $t(1-t) = (1-t)^2 t + t^2(1-t) \in \mathcal{M}(\{t, 1-t\})$, a result of Fekete [PS98, Problem VI.46] (see [Mar08, Proposition 2.7.3] for a modern treatment) implies $\frac{\varepsilon}{2} - s_2 \in \mathcal{M}(\{t, 1-t\})$. \square

Although the tracial version of the Helton-McCullough Positivstellensatz [HM04] fails, we have the following positivity certificate for non-pure trace polynomials.

Corollary 4.8. *Let $\mathcal{M}^{\text{cyc}} \subseteq \text{Sym } \mathbb{T}$ be an archimedean cyclic quadratic module and $a \in \text{Sym } \mathbb{T}$. The following are equivalent:*

- (i) $a(\underline{X}) \succeq 0$ for all $\underline{X} \in \mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{vN}$;
- (ii) $a(\underline{X}) \succeq 0$ for all $\underline{X} \in \mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{\text{II}_1}$;
- (iii) for every $\varepsilon > 0$, there exist sums of (two) squares $s_1, s_2 \in \mathbb{R}[t]$ such that

$$(4.3) \quad a = s_1(a) - s_2(a), \quad \varepsilon - \text{tr}(s_2(a)) \in \mathcal{M}^{\text{cyc}};$$

(iv) for every $\varepsilon > 0$, there exist sums of (two) squares $s_1, s_2 \in \mathbb{R}[t]$ and $q \in \mathcal{M}^{\text{cyc}}$ such that

$$(4.4) \quad \text{tr}(ay) + \varepsilon = \text{tr}(s_1(a)y + s_2(a)(1-y)) + q$$

where y is an auxiliary symmetric free variable. That is, $\text{tr}(ay) + \varepsilon$ is in the cyclic quadratic module generated by $\mathcal{M}^{\text{cyc}}, y, 1-y$ (inside the free trace ring generated by \underline{x}, y).

Proof. (ii) \Leftrightarrow (i) holds by Proposition 2.2. To prove (iii) \Rightarrow (iv), note that (4.3) implies (after multiplication by y and taking the trace) that $\text{tr}(ay) + \varepsilon = \text{tr}(s_1(a)y + s_2(a)(1-y)) + \varepsilon - \text{tr}(s_2(a))$. The implication follows by taking $q = \varepsilon - \text{tr}(s_2(a)) \in \mathcal{M}^{\text{cyc}}$. (iv) \Rightarrow (i) Let (\mathcal{F}, τ) be a tracial pair. If $\text{tr}(ay) + \varepsilon$ belongs to the cyclic quadratic module generated by $\mathcal{M}^{\text{cyc}}, y, 1-y$ for every $\varepsilon > 0$, then $\text{tr}(a(\underline{X})Y) \geq 0$ for all $\underline{X} \in \mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{\mathcal{F}, \tau}$ and positive semidefinite contractions $Y \in \mathcal{F}$. Therefore $a(\underline{X}) \succeq 0$ for all $\underline{X} \in \mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{\mathcal{F}, \tau}$ by Proposition 4.3.

(i) \Rightarrow (iii) Since \mathcal{M}^{cyc} is archimedean, there exists $N > 0$ such that $N - a \in \mathcal{M}^{\text{cyc}}$. After rescaling a we can without loss of generality assume that $1 - a \in \mathcal{M}^{\text{cyc}}$. Suppose that (i) holds. Given an arbitrary $\varepsilon > 0$ let $s_1, s_2 \in \mathbb{R}[t]$ be sums of squares as in Lemma 4.7. Then there are sums of squares $s_3, s_4, s_5 \in \mathbb{R}[t]$ such that

$$s_1 - s_2 = t, \quad \frac{\varepsilon}{2} - s_2 = s_3 + s_4 t + s_5(1-t).$$

By Proposition 4.3 and Theorem 4.4 we have $\text{tr}(s_4(a)a) + \frac{\varepsilon}{2} \in \mathcal{M}^{\text{cyc}}$. Hence

$$(s_3(a) + \text{tr}(s_4(a)a) + \frac{\varepsilon}{2}) + s_5(a)(1-a) \in \mathcal{M}^{\text{cyc}}$$

and therefore

$$a = s_1(a) - s_2(a), \quad \varepsilon - \text{tr}(s_2(a)) \in \mathcal{M}^{\text{cyc}}. \quad \square$$

5. SDP HIERARCHY FOR TRACE OPTIMIZATION

In this section we apply Theorem 3.5 to optimization of pure trace objective functions subject to (pure) trace constraints and a norm boundedness condition. Doing so, we obtain a converging hierarchy of SDP relaxations in Section 5.1. When flatness occurs in this hierarchy, one can extract a finite-dimensional minimizer as shown in Section 5.2. Finally, we apply Proposition 4.3 to handle the more general case of trace polynomials subject to trace constraints and a norm boundedness condition in Section 5.3.

We define the set of *tracial words* (abbreviated as \mathbb{T} -words) by $\{\prod_i \text{tr}(u_i)v \mid u_i, v \in \langle \underline{x} \rangle\}$, which is a subset of \mathbb{T} . The set of *pure trace words* (abbreviated as \mathbb{T} -words) is the subset of \mathbb{T} -words belonging to \mathbb{T} . For instance, $\text{tr}(x_1)^2$ is a \mathbb{T} -word and $\text{tr}(x_1)x_1$ is a \mathbb{T} -word. For $u_i, v \in \langle \underline{x} \rangle$, we define the *tracial degree* of $\prod_i \text{tr}(u_i)v$ as the sum of the degrees of the u_i and the degree of v . The tracial degree of a trace polynomial $f \in \mathbb{T}$ is the length of the longest tracial word involved in f up to cyclic equivalence. Let us denote by $\mathbf{W}_d^{\mathbb{T}}$ (resp. $\mathbf{W}_d^{\mathbb{T}}$) the vector of all \mathbb{T} -words (resp. \mathbb{T} -words) of tracial degree at most d w.r.t. to the degree lexicographic order. Finally, let \mathbb{T}_d (resp. \mathbb{T}_d) denote the span of entries of $\mathbf{W}_d^{\mathbb{T}}$ (resp. $\mathbf{W}_d^{\mathbb{T}}$) in \mathbb{T} (resp. \mathbb{T}), and let $\sigma^{\mathbb{T}}(n, d)$ (resp. $\sigma^{\mathbb{T}}(n, d)$) the dimension of \mathbb{T}_d (resp. \mathbb{T}_d), that is, the length of $\mathbf{W}_d^{\mathbb{T}}$ (resp. $\mathbf{W}_d^{\mathbb{T}}$).

Precise values of $\sigma^{\mathbb{T}}(n, d) \leq \sigma^{\mathbb{T}}(n, d)$ are related to bracelet counting in combinatorics. To get a crude estimate, notice that the number of tracial words of degree d is at least n^d and at most $n^d \cdot 2^d$. Thus

$$\frac{n^{d+1} - 1}{n - 1} \leq \sigma^{\mathbb{T}}(n, d) \leq \frac{(2n)^{d+1} - 1}{2n - 1}.$$

We introduce the notion of trace Hankel and (pure) trace localizing matrices, which can be viewed as tracial analogs of the noncommutative localizing and Hankel matrices (see e.g. [BKP16, Lemma 1.44]). Given $s \in \mathbb{T}$, let us denote $d_s := \lceil \deg s / 2 \rceil$. To s and a linear functional $L : \mathbb{T}_{2d} \rightarrow \mathbb{R}$, one associates the following three matrices:

- (a) the *tracial Hankel matrix* $\mathbf{M}_d^{\mathbb{T}}(L)$ is the symmetric matrix of size $\sigma^{\mathbb{T}}(n, d)$, indexed by \mathbb{T} -words $u, v \in \mathbb{T}_d$, with $(\mathbf{M}_d^{\mathbb{T}}(L))_{u,v} = L(\text{tr}(u^*v))$;
- (b) if $s \in \mathbb{T}$, then the *pure trace localizing matrix* $\mathbf{M}_{d-d_s}^{\mathbb{T}}(sL)$ is the symmetric matrix of size $\sigma^{\mathbb{T}}(n, d - d_s)$, indexed by \mathbb{T} -words $u, v \in \mathbb{T}_{d-d_s}$, with $(\mathbf{M}_{d-d_s}^{\mathbb{T}}(sL))_{u,v} = L(uvs)$;
- (c) the *trace localizing matrix* $\mathbf{M}_{d-d_s}^{\mathbb{T}}(sL)$ is the symmetric matrix of size $\sigma^{\mathbb{T}}(n, d - d_s)$, indexed by \mathbb{T} -words $u, v \in \mathbb{T}_{d-d_s}$, with $(\mathbf{M}_{d-d_s}^{\mathbb{T}}(sL))_{u,v} = L(\text{tr}(u^*sv))$.

Definition 5.1. A matrix \mathbf{M} indexed by \mathbb{T} -words of degree $\leq d$ satisfies the *tracial Hankel condition* if and only if

$$(5.1) \quad \mathbf{M}_{u,v} = \mathbf{M}_{w,z} \text{ whenever } \text{tr}(u^*v) = \text{tr}(w^*z).$$

Remark 5.2. Linear functionals on \mathbb{T}_{2d} and matrices from $\mathbb{S}_{\sigma^{\mathbb{T}}(n,d)}$ satisfying the tracial Hankel condition (5.1) are in bijective correspondence. To a linear functional $L : \mathbb{T}_{2d} \rightarrow \mathbb{R}$, one can assign the matrix $\mathbf{M}_d^{\mathbb{T}}(L)$, defined by $(\mathbf{M}_d^{\mathbb{T}}(L))_{u,v} = L(\text{tr}(u^*v))$, satisfying the tracial Hankel condition, and vice versa.

One can relate the positivity of L and the positive semidefiniteness of its tracial Hankel matrix $\mathbf{M}_d^{\mathbb{T}}(L)$. The proof of the following lemma is straightforward and analogous to its free counterpart [BKP16, Lemma 1.44].

Lemma 5.3. Given a linear functional $L : \mathbb{T}_{2d} \rightarrow \mathbb{R}$, one has $L(\text{tr}(f^*f)) \geq 0$ for all $f \in \mathbb{T}_d$, if and only if, $\mathbf{M}_d^{\mathbb{T}}(L) \succeq 0$. Given $s \in \mathbb{T}$, one has $L(a^2s) \geq 0$ for all $a \in \mathbb{T}_{d-d_s}$, if and only if, $\mathbf{M}_{d-d_s}^{\mathbb{T}}(sL) \succeq 0$. Given $s \in \mathbb{T}$, one has $L(\text{tr}(f^*sf)) \geq 0$ for all $f \in \mathbb{T}_{d-d_s}$, if and only if, $\mathbf{M}_{d-d_s}^{\mathbb{T}}(sL) \succeq 0$.

5.1. SDP hierarchy for pure trace polynomial optimization. For a finite $S \subseteq \mathbb{T}$, $N > 0$ and $d \in \mathbb{N}$ define

$$(5.2) \quad \mathcal{M}(S(N))_d := \left\{ \sum_{i=1}^K a_i^2 s_i \mid K \in \mathbb{N}, a_i \in \mathbb{T}, s_i \in S(N), \deg(a_i^2 s_i) \leq 2d \right\}.$$

Given $b \in \mathbb{T}$ and $p \in \mathbb{R}\langle \underline{x} \rangle$, note that $b^2 \text{tr}(pp^*) = \text{tr}((bp)(bp)^*)$. Therefore, elements of $\mathcal{M}(S(N))_d$ correspond to sums of elements of the form

$$(5.3) \quad a_1^2 s, \quad a_2^2 (N^k - \text{tr}(x_j^{2k})), \quad \text{tr}(ff^*),$$

which are of degree at most $2d$, for $a_i \in \mathbb{T}$, $s \in S$, $1 \leq j \leq n$, $k \in \mathbb{N}$, $f \in \mathbb{T}$.

Given a pure trace polynomial $a \in \mathbb{T}$, one can then use $\mathcal{M}(S(N))_d$ for $d = 1, 2, \dots$ to design a hierarchy of semidefinite relaxations for minimizing $a \in \mathbb{T}$ over the von Neumann semialgebraic sets $\mathcal{D}_{S[N]}^{\text{vN}}$ or $\mathcal{D}_{S[N]}^{\text{II}_1}$.

Let us define a_{\min} and $a_{\min}^{\text{II}_1}$ as follows:

$$(5.4) \quad a_{\min} := \inf\{a(\underline{A}) \mid \underline{A} \in \mathcal{D}_{S[N]}\},$$

$$(5.5) \quad a_{\min}^{\text{II}_1} := \inf\{a(\underline{A}) \mid \underline{A} \in \mathcal{D}_{S[N]}^{\text{II}_1}\} = \inf\{a(\underline{A}) \mid \underline{A} \in \mathcal{D}_{S[N]}^{\text{vN}}\}.$$

Here the equality in (5.5) holds by Proposition 2.2. Since $\mathcal{D}_{S[N]}$ is a subset of $\mathcal{D}_{S[N]}^{\text{vN}}$, one has $a_{\min}^{\text{II}_1} \leq a_{\min}$. Let $d_{\min} := \max\{d_s : s \in \{a\} \cup S(N)\}$. Then, one can under-approximate $a_{\min}^{\text{II}_1}$ via the following hierarchy of SDP programs, indexed by $d \geq d_{\min}$:

$$(5.6) \quad a_{\min,d} = \sup\{m \mid a - m \in \mathcal{M}(S(N))_d\}.$$

Lemma 5.4. *The dual of (5.6) is the following SDP problem:*

$$(5.7) \quad \begin{aligned} & \inf_{\substack{L: \mathbb{T}_{2d} \rightarrow \mathbb{R} \\ L \text{ linear}}} L(a) \\ \text{s.t. } & (\mathbf{M}_d^{\mathbb{T}}(L))_{u,v} = (\mathbf{M}_d^{\mathbb{T}}(L))_{w,z}, \quad \text{whenever } \text{tr}(u^*v) = \text{tr}(w^*z), \\ & (\mathbf{M}_d^{\mathbb{T}}(L))_{1,1} = 1, \\ & \mathbf{M}_d^{\mathbb{T}}(L) \succeq 0, \\ & \mathbf{M}_{d-d_s}^{\mathbb{T}}(sL) \succeq 0, \quad \text{for all } s \in S, \\ & \mathbf{M}_{d-k}^{\mathbb{T}}((N^k - \text{tr}(x_j^{2k}))L) \succeq 0, \quad \text{for all } j = 1, \dots, n, k \leq d. \end{aligned}$$

Proof. Let us denote by $(\mathcal{M}(S(N))_d)^\vee$ the dual space of $\mathcal{M}(S(N))_d$. From (5.3), one has

$$\begin{aligned} (\mathcal{M}(S(N))_d)^\vee &= \{L : \mathbb{T}_{2d} \rightarrow \mathbb{R} \mid L \text{ linear}, \\ & L(a^2s) \geq 0, \forall s \in S, \forall a \in \mathbb{T}_{d-d_s}, \\ & L(a^2(N^k - \text{tr}(x_j^{2k}))) \geq 0, \forall j = 1, \dots, n, \forall k \leq d, \forall a \in \mathbb{T}_{d-k}, \\ & L(\text{tr}(ff^*)) \geq 0, \forall f \in \mathbb{T}_d\} \end{aligned}$$

By using a standard Lagrange duality approach, we obtain the dual of SDP (5.6):

$$(5.8) \quad a_{\min,d} = \sup_{a-m \in \mathcal{M}(S(N))_d} m = \sup_m \inf_{L \in (\mathcal{M}(S(N))_d)^\vee} (m + L(a - m))$$

$$(5.9) \quad \leq \inf_{L \in (\mathcal{M}(S(N))_d)^\vee} \sup_m (m + L(a - m))$$

$$(5.10) \quad = \inf_{L \in (\mathcal{M}(S(N))_d)^\vee} (L(a) + \sup_m m(1 - L(1)))$$

$$(5.11) \quad = \inf_L \{L(f) \mid L \in (\mathcal{M}(S(N))_d)^\vee, L(1) = 1\},$$

The second equality in (5.8) comes from the fact that the inner minimization problem gives minimal value 0 if and only if $a - m \in \mathcal{M}(S(N))_d$. The inequality in (5.9) trivially holds. The inner maximization problem in (5.10) is bounded with maximum value 0 if and only if $L(1) = 1$. Eventually, (5.11) is equivalent to SDP (5.7) by Remark 5.2 and Lemma 5.3. \square

Before proving that SDP (5.6) satisfies strong duality, we recall that an ε -neighborhood of 0 is the set \mathcal{N}_ε defined for a given $\varepsilon > 0$ by:

$$\mathcal{N}_\varepsilon := \bigcup_{k \in \mathbb{N}} \left\{ \underline{A} := (A_1, \dots, A_n) \in \mathbb{S}_k^n : \varepsilon^2 - \sum_{i=1}^n A_i^2 \succeq 0 \right\}.$$

Lemma 5.5. *If $f \in \mathbb{T}$ vanishes on an ε -neighborhood of 0, then $f = 0$.*

Proof. We rely on the standard multilinearization trick and the fact that a trace polynomial $f \in \mathbb{T}_d$ cannot be a trace identity on $(d+1) \times (d+1)$ matrices, as a consequence of [Pro76, Theorem 4.5 (b)]. Since f vanishes on all n -tuples of $(d+1) \times (d+1)$ matrices $\underline{A} \in \mathcal{N}_\varepsilon$, one has $f = 0$. \square

Theorem 5.6. *Let $S[N]$ be as in (3.3) and suppose that \mathcal{D}_S contains an ε -neighborhood of 0. Then SDP (5.6) satisfies strong duality, i.e., there is no duality gap between SDP (5.7) and SDP (5.6).*

Proof. The strong duality statement is proved as in [CKP12, Proposition 4.4]. For this, we construct a linear map $L : \mathbb{T}_{2d} \rightarrow \mathbb{R}$ which is a strictly feasible solution of SDP (5.7), namely $L(1) = 1$, $L(a^2s) > 0$ for all $s \in S$ and for all nonzero $a \in \mathbb{T}_{d-d_s}$, $L(a^2(N^k - \text{tr}(x_j^{2k}))) > 0$ for all $j, k \leq d$ and for all nonzero $a \in \mathbb{T}_{d-k}$, and $L(\text{tr}(ff^*)) > 0$ for all nonzero $f \in \mathbb{T}_d$. Let us pick $m > d$ and consider the set \mathcal{U} of $m \times m$ matrices from $\mathcal{D}_{S[N]}$ with rational entries, written as

$$\mathcal{U} = \{ \underline{A}^{(k)} \mid k \in \mathbb{N}, \underline{A}^{(k)} \in \mathcal{D}_{S[N]}^m \}.$$

Note that \mathcal{U} contains a dense subset of $m \times m$ matrices in \mathcal{N}_ε . Let us define

$$L := \sum_{k=1}^{\infty} 2^{-k} \frac{L_{\underline{A}^{(k)}}}{\|L_{\underline{A}^{(k)}}\|},$$

with

$$L_{\underline{A}} : \mathbb{T}_{2d} \rightarrow \mathbb{R}, \quad f \mapsto \text{tr} f(\underline{A}),$$

for all $\underline{A} \in \mathcal{U}$. This functional L is obviously linear and unital. One has $L(\text{tr}(ff^*)) \geq 0$ for all $f \in \mathbb{T}_d$. Now let us assume that $L(\text{tr}(ff^*)) = 0$ for some $f \in \mathbb{T}_d$. This implies that for all $k \in \mathbb{N}$ one has $\text{tr}(f(\underline{A}^{(k)})f^*(\underline{A}^{(k)})) = 0$, thus $f(\underline{A}^{(k)})f^*(\underline{A}^{(k)}) = 0$, which in turn implies that $f(\underline{A}^{(k)}) = 0$. By density of \mathcal{U} in $\mathcal{N}_\varepsilon \cap \mathbb{S}_m^n$, $f(\underline{A}) = 0$ for all $\underline{A} \in \mathcal{N}_\varepsilon \cap \mathbb{S}_m^n$. As m was arbitrary, f vanishes on \mathcal{N}_ε . By Lemma 5.5, one has $f = 0$. The two other positivity conditions are proved in a similar fashion. \square

Corollary 5.7. *The hierarchy of SDP programs (5.6) provides a sequence of lower bounds $(a_{\min,d})_{d \geq d_{\min}}$ monotonically converging to $a_{\min}^{\text{II}_1}$.*

Proof. As $\mathcal{M}(S(N))_d \subseteq \mathcal{M}(S(N))_{d+1}$, one has $a_{\min,d} \leq a_{\min,d+1}$. Furthermore, Theorem 3.5 implies that for each $m \in \mathbb{N}$, there exists $d(m) \in \mathbb{N}$ such that $a - a_{\min}^{\text{II}_1} + \frac{1}{m} \in \mathcal{M}(S(N))_{d(m)}$. Thus one has

$$a_{\min}^{\text{II}_1} - \frac{1}{m} \leq a_{\min,d(m)},$$

which implies that

$$\lim_{d \rightarrow \infty} a_{\min,d} = a_{\min}^{\text{II}_1}. \quad \square$$

5.2. Finite-dimensional GNS representations and minimizer extraction.

The goal of this section is to derive an algorithm to extract minimizers of pure trace polynomial optimization problems. The forthcoming statements can be seen as “pure trace” variants of the results derived in the context of commutative polynomials [CF98], eigenvalue optimization of noncommutative polynomials [McC01, Lemma 2.2] (see also [PNA10], [AL12, Chapter 21] and [BKP16, Theorem 1.69]), and trace optimization of noncommutative polynomials [BCKP13].

Definition 5.8. *Suppose $L : \mathbb{T}_{2d+2\delta} \rightarrow \mathbb{R}$ is a tracial linear functional with restriction $\tilde{L} : \mathbb{T}_{2d} \rightarrow \mathbb{R}$. We associate to L and \tilde{L} the Hankel matrices $\mathbf{M}_{d+\delta}^{\mathbb{T}}(L)$ and $\mathbf{M}_d^{\mathbb{T}}(\tilde{L})$ respectively, and get the block form*

$$\mathbf{M}_{d+\delta}^{\mathbb{T}}(L) = \begin{bmatrix} \mathbf{M}_d^{\mathbb{T}}(\tilde{L}) & B \\ B^T & C \end{bmatrix}.$$

We say that L is δ -flat or that L is a δ -flat extension of \tilde{L} , if $\mathbf{M}_{d+\delta}^{\mathbb{T}}(L)$ is flat over $\mathbf{M}_d^{\mathbb{T}}(\tilde{L})$, i.e., if $\text{rank } \mathbf{M}_{d+\delta}^{\mathbb{T}}(L) = \text{rank } \mathbf{M}_d^{\mathbb{T}}(\tilde{L})$.

Suppose L is δ -flat and let $r := \text{rank } \mathbf{M}_d^{\mathbb{T}}(L) = \text{rank } \mathbf{M}_{d+\delta}^{\mathbb{T}}(L)$. Since $\mathbf{M}_{d+\delta}^{\mathbb{T}}(L) \succeq 0$, we obtain the Gram matrix decomposition $\mathbf{M}_{d+\delta}^{\mathbb{T}}(L) = [(\mathbf{u}, \mathbf{w})]_{u,w}$ with vectors $\mathbf{u}, \mathbf{w} \in \mathbb{R}^r$, where the labels u, v are \mathbb{T} -words of degree at most $d + \delta$. Then, we define the following finite-dimensional Hilbert space

$$\mathcal{H} := \text{span} \{ \mathbf{w} \mid \deg w \leq d + \delta \} = \text{span} \{ \mathbf{w} \mid \deg w \leq d \},$$

where the equality is a consequence of the flatness assumption. Afterwards, one can follow the steps performed in Appendix A (see (A.4)) and consider, for each $p \in \mathbb{T}$, the multiplication operator $\hat{\chi}_p$ on \mathcal{H} and the \star -representation $\pi : \mathbb{T} \rightarrow \mathcal{B}(\mathcal{H})$ defined by $\pi(p) = \hat{\chi}_p$. Let \mathbf{v} be the vector representing 1 in \mathcal{H} ; then $L(p) = \langle \pi(p)\mathbf{v}, \mathbf{v} \rangle$ for all $p \in \mathbb{T}$. In general, elements of $\pi(\mathbb{T})$ are central in $\pi(\mathbb{T})$; if they are actually scalar multiples of the identity on \mathcal{H} , then π is not just a \star -representation, but it respects trace in the sense that $\pi(f) = f(\pi(x_1), \dots, \pi(x_n))$ for every $f \in \mathbb{T}$. This fact applies to our SDP hierarchy as follows.

Proposition 5.9. *Given $S \cup \{a\} \subseteq \mathbb{T}_{2d}$, let $S[N]$ be as in (3.3). Set $\delta := \max\{\lfloor \deg s/2 \rfloor : s \in S[N]\}$. Assume that L is a δ -flat optimal solution of SDP (5.7), and assume that $\pi(\mathbb{T}) = \mathbb{R}$, where $\pi : \mathbb{T} \rightarrow \mathcal{B}(\mathcal{H})$ is the \star -representation constructed above. Then, one has*

$$(5.12) \quad a_{\min, d+\delta} = L(a) = a_{\min}^{\text{II}_1}.$$

Moreover, there are finitely many n -tuples $\underline{A}^{(j)}$ of symmetric matrices, and positive scalars λ_j with $\sum_j \lambda_j = 1$, such that $a_{\min}^{\text{II}_1} = a(\bigoplus_j \underline{A}^{(j)})$, where the tracial state is given by

$$w \left(\bigoplus_j \underline{A}^{(j)} \right) \mapsto \sum_j \lambda_j \text{tr}(w(\underline{A}^{(j)}))$$

for $w \in \langle \underline{x} \rangle$.

Proof. For $i = 1, \dots, n$ let $A_i = \hat{\chi}_{x_i}$ be the left multiplication by x_i on \mathcal{H} , i.e., for each \mathbb{T} -word $w \in \mathbb{T}_d$, $A_i w$ is the vector from \mathcal{H} corresponding to the label $x_i w$. The operators A_i are well-defined (thanks to the flatness assumption) and symmetric. After choosing an orthonormal basis of \mathcal{H} we can view A_i as $r \times r$ symmetric matrices. Let $\underline{A} := (A_1, \dots, A_n)$, and let $\mathcal{A} \subseteq \mathbb{M}_r$ be the algebra generated by A_1, \dots, A_n . Since $\pi(\mathbb{T}) = \mathbb{R}$, the map $\tau : \mathcal{A} \rightarrow \mathbb{R}$ given by $q(\underline{A}) \mapsto \pi(\text{tr } q) = L(q)$ for $q \in \mathbb{R}\langle \underline{x} \rangle$ is a well-defined faithful tracial state on \mathcal{A} . For each $s \in S$, one has $s(\underline{A}) = \langle \pi(s)\mathbf{v}, \mathbf{v} \rangle = L(s) \geq 0$, where the last inequality follows from the fact that $\mathbf{M}_{d-d_s}^{\mathbb{T}}(sL) \succeq 0$ as L is a feasible solution of SDP (5.7). Similarly, one has $N^k - \tau(A_j^{2k}) \geq 0$, for all $j = 1, \dots, n, k \leq d$. Therefore, $\underline{A} \in \mathcal{D}_{S[N]}^{vN}$.

Eventually, $a_{\min, d+\delta} = L(a) \leq a_{\min}^{\text{II}_1}$, where the first equality is the strong duality statement from Theorem 5.6. In addition, one has $a_{\min}^{\text{II}_1} \leq a(\underline{A}) = L(a)$, yielding the desired result (5.12).

We get a tracial representation of the optimizer for $a_{\min}^{\text{II}_1}$ by performing the Artin-Wedderburn block diagonalization on the algebra \mathcal{A} . This step relies on the Wedderburn theorem [Lam13, Chapter 1]. By [BKP16, Proposition 1.68], there are finitely many tuples of symmetric matrices $\underline{A}^{(j)}$ and positive scalars λ_j with $\sum_j \lambda_j = 1$ such that

$$\tau(q(\underline{A})) = \sum_j \lambda_j \text{tr}(q(\underline{A}^{(j)}))$$

for all $q \in \mathbb{R}\langle \underline{x} \rangle$. □

Remark 5.10. *The condition $\pi(\mathbb{T}) = \mathbb{R}$ in Proposition 5.9 in particular holds if L is an extreme optimal solution of (5.7) (cf. the last part of the proof in Appendix A). In practice modern SDP solvers rely on interior-point methods using the so-called “self-dual embedding” technique [WSV12, Chapter 5]. Therefore, they will always converge towards an optimum solution of maximum rank, yielding an extreme linear functional; see [LLR08, §4.4.1] for more details.*

Remark 5.11. *Proposition 5.9 guarantees that in the presence of a flat extension, there is an optimizer for $a_{\min}^{\text{II}_1}$ arising from a finite-dimensional tracial pair (\mathcal{F}, τ) ; furthermore, the dimensions of $\underline{A}^{(j)}$ and the scalars λ_j explicitly determine \mathcal{F} and τ , respectively. It is sensible to ask whether $a_{\min} = a_{\min}^{\text{II}_1}$, that is, whether the optimum can be approximated arbitrarily well with a finite-dimensional factor, i.e., from $\mathcal{D}_{S[N]}$. If there exist sequences of positive rational numbers $(\lambda_j^{(m)})_m$ such that $\sum_j \lambda_j^{(m)} = 1$ for all $m \in \mathbb{N}$, $\lim_m \lambda_j^{(m)} = \lambda_j$ for all j , and $\bigoplus_j \underline{A}^{(j)} \in \mathcal{D}_{S[N]}^{vN}$ whenever the tracial state is given by*

$$(5.13) \quad w \left(\bigoplus_j \underline{A}^{(j)} \right) \mapsto \sum_j \lambda_j^{(n)} \text{tr}(w(\underline{A}^{(j)})) \quad \text{for } w \in \langle \underline{x} \rangle,$$

then $a_{\min} = a_{\min}^{\text{II}_1}$. Indeed, a finite-dimensional tracial pair with the rational-coefficient tracial state as in (5.13) embeds into a finite-dimensional factor. However, in general $a_{\min} \neq a_{\min}^{\text{II}_1}$ even if \mathcal{D}_S contains an ε -neighborhood of 0 and $a_{\min}^{\text{II}_1}$ admits a finite-dimensional optimizer; see the following example.

Example 5.12. Fix $n = 1$, i.e., $\mathbb{T} = \mathbb{R}[\text{tr}(x_1^i) \mid i \in \mathbb{N}]$. For $k \in \mathbb{N}$ let

$$s_k := 1 + (\sqrt{2} + 1)^2 - \left(\text{tr}((x_1^2 - 2x_1)^2) + (\sqrt{2} - \text{tr}(x_1))^2 \right) \text{tr}(x_1^{2k}) \in \mathbb{T}.$$

Let X be a symmetric matrix. Then $X^2 \neq 2X$ or $\text{tr}(X) \neq \sqrt{2}$. Furthermore, if X is a contraction, then

$$0 \preceq 2X - X^2 \preceq I, \quad |\sqrt{2} - \text{tr}(X)| \leq \sqrt{2} + 1, \quad \text{tr}(X^{2k}) \leq 1 \quad \text{for all } k \in \mathbb{N}.$$

On the other hand, if X is not a contraction, then there is $k \in \mathbb{N}$ such that

$$\text{tr}(X^{2k}) > \frac{1 + (\sqrt{2} + 1)^2}{\text{tr}((X^2 - 2X)^2) + (\sqrt{2} - \text{tr}(X))^2}.$$

Let $S = \{s_k \mid k \in \mathbb{N}\}$ and $a = -\text{tr}(x_1)$. Then $\mathcal{D}_S = \mathcal{D}_{1-x_1^2}$ by the above observations, and consequently $a_{\min} = -1$. On the other hand, consider the tracial pair (\mathbb{R}^2, τ) with $\tau(\xi_1, \xi_2) = \frac{1}{\sqrt{2}}\xi_1 + (1 - \frac{1}{\sqrt{2}})\xi_2$. Then $Y = (2, 0) \in \mathbb{R}^2$ satisfies $Y^2 = 2Y$ and $\tau(Y) = \sqrt{2}$, so $Y \in \mathcal{D}_S^{\mathbb{N}}$. Therefore $a_{\min}^{\text{H}_1} < a(Y) = -\sqrt{2}$.

The proof of Proposition 5.9 gives the following procedure for minimizer extraction.

Algorithm 5.13. *PureTraceGNS*

Input: an extreme δ -flat linear $L : \mathbb{T}_{2d+2\delta} \rightarrow \mathbb{R}$ solution of (5.7).

- 1: Let us consider the set of \mathbb{T} -words $\{w_i\}$ of degree at most $\leq d$, such that \mathcal{C} , the matrix consisting of columns of $\mathbf{M}(L)$ indexed by the words w_1, \dots, w_r , has full rank. Assume $w_1 = 1$.
 - 2: Let $\mathbf{M}(\hat{L})$ be the principal submatrix of $\mathbf{M}(L)$ of columns and rows indexed by w_1, \dots, w_r .
 - 3: Let C be the Cholesky factor of $\mathbf{M}(\hat{L})$, i.e., $C^T C = \mathbf{M}(\hat{L})$.
 - 4: **for** $i \in \{1, \dots, n\}$ **do**
 - 5: Let \mathcal{C}_i be the matrix consisting of columns of $\mathbf{M}(L)$ indexed by $x_i w_1, \dots, x_i w_r$.
 - 6: Compute \bar{A}_i as a solution of the system $\mathcal{C} \bar{A}_i = \mathcal{C}_i$.
 - 7: Let $A_i = C \bar{A}_i C^{-1}$.
 - 8: **end for**
 - 9: Compute $\mathbf{v} = C e_1$. $\triangleright e_1 = (1, 0, \dots, 0)$
 - 10: Let $\mathcal{A} \subseteq \mathbb{M}_r$ be the algebra generated by A_1, \dots, A_n . Compute an orthogonal matrix Q performing the simultaneous block-diagonalization of A_1, \dots, A_n by [MKKK10, Algorithm 4.1]. $\triangleright Q^T A Q = \{\text{Diag}(B^{(1)}, \dots, B^{(k)}) \mid B^{(i)} \in \mathcal{A}_i\}$
where $\mathcal{A}_1, \dots, \mathcal{A}_k$ are simple \star -algebras over \mathbb{R}
 - 11: Compute $Q^T A_i Q = \text{Diag}(A_i^{(1)}, \dots, A_i^{(k)})$ for each $i = 1, \dots, n$, and $Q^T \mathbf{v} = ((\mathbf{v}^1)^T, \dots, (\mathbf{v}^k)^T)^T$.
 - 12: Compute $\lambda_j = \|\mathbf{v}^j\|$, and $\underline{A}^{(j)} = (A_1^{(j)}, \dots, A_n^{(j)})$, for all $j = 1, \dots, k$.
- Output:** $(\underline{A}^{(1)}, \dots, \underline{A}^{(k)})$ and $(\lambda_1, \dots, \lambda_k)$.

The correctness of the procedure *PureTraceGNS* follows from the proof of Proposition 5.9.

Corollary 5.14. *The procedure *PureTraceGNS* described in Algorithm 5.13 is sound and returns the n -tuples $\underline{A}^{(j)}$ and λ_j from Proposition 5.9.*

Remark 5.15. *Note that when flatness occurs, Proposition 5.9 guarantees convergence (actually stabilization) of our SDP hierarchy even if there is no ε -neighborhood of 0 in the feasible set. Moreover, while a flat extension is evasive from a numerical point of view, an “almost” flat extension, which is a much more viable output of an SDP solver, is likely sufficient [KPV18].*

5.3. SDP hierarchy for trace polynomial optimization. Here we describe the reduction from the general trace setting to the pure trace setting.

Let $S \subseteq \text{Sym } \mathbb{T}$ and $N > 0$. Denote

$$(5.14) \quad \tilde{S} = \{\text{tr}(fsf^*) \mid s \in S, f \in \mathbb{T}\} \subseteq T.$$

Proposition 5.16. *Let $S \subseteq \text{Sym } \mathbb{T}$, $N > 0$, and let \tilde{S} be as in (5.14). Then $\mathcal{D}_{\tilde{S}[N]}^{\mathcal{F}, \tau} = \mathcal{D}_{S[N]}^{\mathcal{F}, \tau}$ for any tracial pair (\mathcal{F}, τ) . Furthermore, the following are equivalent for $a \in T$:*

- (i) $a(\underline{X}) \geq 0$ for all $\underline{X} \in \mathcal{D}_{S[N]}^{vN}$;
- (ii) $a(\underline{X}) \geq 0$ for all $\underline{X} \in \mathcal{D}_{S[N]}^{\text{II}_1}$;
- (iii) $a + \varepsilon \in \mathcal{M}(\tilde{S}(N))$ for all $\varepsilon > 0$.

Proof. The equality $\mathcal{D}_{\tilde{S}[N]}^{\mathcal{F}, \tau} = \mathcal{D}_{S[N]}^{\mathcal{F}, \tau}$ follows from Proposition 4.3. Consequently, (i) \Leftrightarrow (ii) \Leftrightarrow (iii) holds by Theorem 3.5. \square

For all $d \in \mathbb{N}$, one has

$$\mathcal{M}(\tilde{S}(N))_d = \left\{ \sum_{i=1}^K a_i^2 s_i \mid K \in \mathbb{N}, a_i \in \mathbb{T}, s_i \in \tilde{S}(N), \deg(a_i^2 s_i) \leq 2d \right\}.$$

Therefore, elements of $\mathcal{M}(\tilde{S}(N))_d$ corresponds to sums of elements of the form

$$(5.15) \quad \text{tr}(f_1 s f_1^*), \quad a^2(N^k - \text{tr}(x_j^{2k})), \quad \text{tr}(f_2 f_2^*),$$

which are of degree at most $2d$, for $f_i \in \mathbb{T}$, $a \in \mathbb{T}$, $s \in S$, $1 \leq j \leq n$, $k \in \mathbb{N}$.

As in Section 5.1, given $a \in \mathbb{T}$, one can under-approximate $a_{\min}^{\text{II}_1}$ via the following hierarchy of SDP programs, indexed by $d \geq d_{\min}$:

$$(5.16) \quad \tilde{a}_{\min, d} = \sup\{m \mid a - m \in \mathcal{M}(\tilde{S}(N))_d\}.$$

The dual of (5.16) is obtained by replacing the pure trace localizing matrix constraints in SDP (5.7) by trace localizing matrix constraints associated to each $s \in S$:

$$(5.17) \quad \begin{aligned} & \inf_{\substack{L: \mathbb{T}_{2d} \rightarrow \mathbb{R} \\ L \text{ linear}}} L(a) \\ \text{s.t.} \quad & (\mathbf{M}_d^{\mathbb{T}}(L))_{u,v} = (\mathbf{M}_d^{\mathbb{T}}(L))_{w,z}, \quad \text{whenever } \text{tr}(u^* v) = \text{tr}(w^* z), \\ & (\mathbf{M}_d^{\mathbb{T}}(L))_{1,1} = 1, \\ & \mathbf{M}_d^{\mathbb{T}}(L) \succeq 0, \\ & \mathbf{M}_{d-d_s}^{\mathbb{T}}(sL) \succeq 0, \quad \text{for all } s \in S, \\ & \mathbf{M}_{d-k}^{\mathbb{T}}((N^k - \text{tr}(x_j^{2k}))L) \succeq 0, \quad \text{for all } j = 1, \dots, n, k \leq d. \end{aligned}$$

As in Theorem 5.6, one can prove that if \mathcal{D}_S contains an ε -neighborhood of 0, then there is no duality gap between SDP (5.17) and SDP (5.16). In addition, the hierarchy of SDP programs (5.16) provides a sequence of lower bounds monotonically converging to $a_{\min}^{\Pi_1}$.

Remark 5.17. *There are several variations of (5.17) that lead to $a_{\min}^{\Pi_1}$. For example, one can replace the d matrix inequalities $\mathbf{M}_{d-k}^T((N^k - \text{tr}(x_j^{2k}))L) \succeq 0$ (for a fixed j) with a single matrix inequality $\mathbf{M}_{d-1}^T((N - x_j^2)L) \succeq 0$ since*

$$N^k - \text{tr}(x_j^{2k}) = \text{tr} \left(\sum_{i=1}^{k-1} (\sqrt{N}^{k-1-i} x_j^i)(N - x_j^2)(\sqrt{N}^{k-1-i} x_j^i) \right).$$

While this modification produces a somewhat simpler-looking SDP, note that the new matrix constraint has size $\sigma^{\mathbb{T}}(n, d-1)$, while the combined size of the replaced constraints equals $\sum_{k=1}^d \sigma^{\mathbb{T}}(n, k)$, which is less than $\sigma^{\mathbb{T}}(n, d-1)$.

In practice, when given a concrete set of constraints S , one should attempt to simplify (5.17) before solving it, since its most general form can contain superfluous inequalities with respect to S .

Finally, the next result provides an alternative characterization of (not necessarily pure) trace polynomials positive on tracial semialgebraic sets (cf. Corollary 4.8).

Proposition 5.18. *Let $S \subseteq \text{Sym } \mathbb{T}$, $N > 0$, and let \tilde{S} be as in (5.14). For $a \in \text{Sym } \mathbb{T}$, the following are equivalent:*

- (i) $a(\underline{X}) \succeq 0$ for all $\underline{X} \in \mathcal{D}_{S[N]}^{vN}$;
 - (ii) $a(\underline{X}) \succeq 0$ for all $\underline{X} \in \mathcal{D}_{S[N]}^{\Pi_1}$;
 - (iii) for every $\varepsilon > 0$, there exist sums of (two) squares $s_1, s_2 \in \mathbb{R}[t]$ such that
- $$(5.18) \quad a = s_1(a) - s_2(a), \quad \varepsilon - \text{tr}(s_2(a)) \in \mathcal{M}(\tilde{S}(N));$$
- (iv) for every $\varepsilon > 0$, there exist sums of (two) squares $s_1, s_2 \in \mathbb{R}[t]$ and $q \in \mathcal{M}(\tilde{S}(N))$ such that

$$(5.19) \quad \text{tr}(ay) + \varepsilon = \text{tr}(s_1(a)y + s_2(a)(1-y)) + q$$

where y is an auxiliary symmetric free variable.

Proof. (i) \Leftrightarrow (ii) holds by Proposition 2.2, and (iii) \Rightarrow (iv) follows by taking $q = \varepsilon - \text{tr}(s_2(a)) \in \mathcal{M}(\tilde{S}(N))$. Furthermore, (iv) \Rightarrow (i) holds by Propositions 4.3 and 5.16.

(i) \Rightarrow (iii) There is $N' > 0$, dependent on N and a , such that $N' - a \succeq 0$ on $\mathcal{D}_{S[N]}^{vN}$. After rescaling a we can without loss of generality assume that $1 - a \succeq 0$ on $\mathcal{D}_{S[N]}^{vN}$. Suppose that (i) holds. Given an arbitrary $\varepsilon > 0$ let $s_1, s_2 \in \mathbb{R}[t]$ be sums of squares as in Lemma 4.7. Then there are sums of squares $s_3, s_4, s_5 \in \mathbb{R}[t]$ such that

$$s_1 - s_2 = t, \quad \frac{\varepsilon}{2} - s_2 = s_3 + s_4 t + s_5(1-t).$$

By Propositions 4.3 and 5.16 we have

$$\text{tr}(s_3(a) + s_4(a)a + s_5(a)(1-a)) + \frac{\varepsilon}{2} \in \mathcal{M}(\tilde{S}(N)).$$

Therefore

$$a = s_1(a) - s_2(a), \quad \varepsilon - \operatorname{tr}(s_2(a)) \in \mathcal{M}(\tilde{S}(N)). \quad \square$$

Note that Proposition 5.18 allows one to certify that a given trace polynomial is positive semidefinite on a tracial semialgebraic set. Constructing a hierarchy of SDP programs converging to the minimal eigenvalue of trace polynomials is postponed for future work. As Example 4.6 indicates, it cannot be simply derived from our scheme for the pure trace polynomial objective function; namely, the norm of an operator cannot be uniformly estimated with traces in a dimension-free way.

6. EXAMPLES AND APPLICATIONS

In this section we present some experimental results indicating the strength and computational aspects of the SDP hierarchy in Section 5. First we give a toy example of optimizing a pure trace polynomial over all projections in tracial von Neumann algebras (Section 6.1). Next we describe how our algorithms can be used for finding upper bounds on quantum violations of polynomial Bell inequalities in quantum information theory (Section 6.2). While the SDPs were solved using SeDuMi in Matlab, the sparse input matrices were constructed with Mathematica.

6.1. A toy example. Consider the optimization problem

$$(6.1) \quad \begin{aligned} \inf \quad & \tau(X_1 X_2 X_3) + \tau(X_1 X_2) \tau(X_3) \\ \text{s.t.} \quad & X_j^2 = X_j^* = X_j \quad \text{for } j = 1, 2, 3. \end{aligned}$$

over triples (X_1, X_2, X_3) of operators in tracial pairs (\mathcal{F}, τ) . Note that if \mathcal{F} is a commutative von Neumann algebra with a tracial state τ and $X_1, X_2, X_3 \in \mathcal{F}$ are projections, then $\tau(X_1 X_2 X_3), \tau(X_1 X_2), \tau(X_3) \geq 0$. Hence if (6.1) were restricted only to commutative von Neumann algebras, the solution would be 0. On the other hand, the projections

$$X_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} \frac{1}{16} & \frac{\sqrt{15}}{16} \\ \frac{\sqrt{15}}{16} & \frac{15}{16} \end{pmatrix}, \quad X_3 = \begin{pmatrix} \frac{3}{8} & -\frac{\sqrt{15}}{8} \\ -\frac{\sqrt{15}}{8} & \frac{5}{8} \end{pmatrix}$$

give

$$\operatorname{tr}(X_1 X_2 X_3) + \operatorname{tr}(X_1 X_2) \operatorname{tr}(X_3) = -\frac{1}{32}.$$

We next show that $-\frac{1}{32}$ is actually the solution of (6.1).

Let $n = 3$, $a = \operatorname{tr}(x_1 x_2 x_3) + \operatorname{tr}(x_1 x_2) \operatorname{tr}(x_3)$ and $S = \{x_j^2 - x_j, x_j - x_j^2 : j = 1, 2, 3\}$. By Section 5.3, the solution of (6.1) equals $\lim_{n \rightarrow \infty} \check{a}_{\min, d}$, where $\check{a}_{\min, d}$ is the solution of (5.7) for $d \geq 2$. In this particular example, the constraints can be used to vastly simplify (5.7). Namely, it suffices to consider only tracial words without consecutive repetitions of x_j ; furthermore, the last two lines in (5.7) are then superfluous. To state this concretely, let us introduce some auxiliary notation.

A \mathbb{T} -word is *square-reduced* if no proper powers of x_1, x_2, x_3 appear in it. To each \mathbb{T} -word w we can assign the square-reduced \mathbb{T} -word $r(w)$ by repeatedly replacing x_j^2 with x_j . Let \mathbf{W}_d^r be the vector of all square-reduced \mathbb{T} -words of tracial degree at most d , and let R_d be the span of entries of \mathbf{W}_d^r . Given a linear functional

$L : R_{2d} \rightarrow \mathbb{R}$, the *square-reduced tracial Hankel matrix* $\mathbf{M}_d^r(L)$ is indexed by \mathbf{W}_d^r and $(\mathbf{M}_d^r(L))_{u,v} = L(\text{tr}(r(u^*v)))$. Then $\check{a}_{\min,d}$ is the solution of the SDP

$$(6.2) \quad \begin{aligned} & \inf_{\substack{L: R_{2d} \rightarrow \mathbb{R} \\ L \text{ linear}}} L(a) \\ \text{s.t.} \quad & (\mathbf{M}_d^r(L))_{u,v} = (\mathbf{M}_d^r(L))_{w,z}, \quad \text{whenever } \text{tr}(u^*v) = \text{tr}(w^*z), \\ & (\mathbf{M}_d^r(L))_{1,1} = 1, \\ & \mathbf{M}_d^r(L) \succeq 0. \end{aligned}$$

We start with $d = 2$. The matrix $\mathbf{M}_2^r(L)$ is indexed by reduced tracial words

$$\begin{aligned} & 1, x_1, x_2, x_3, \text{tr}(x_1), \text{tr}(x_2), \text{tr}(x_3), \\ & x_1x_2, x_2x_1, x_1x_3, x_3x_1, x_2x_3, x_3x_2, \text{tr}(x_1x_2), \text{tr}(x_1x_3), \text{tr}(x_2x_3), \\ & \text{tr}(x_1)x_1, \text{tr}(x_1)x_2, \text{tr}(x_1)x_3, \text{tr}(x_2)x_1, \text{tr}(x_2)x_2, \text{tr}(x_2)x_3, \text{tr}(x_3)x_1, \text{tr}(x_3)x_2, \text{tr}(x_3)x_3, \\ & \text{tr}(x_1)^2, \text{tr}(x_2)^2, \text{tr}(x_3)^2, \text{tr}(x_1) \text{tr}(x_2), \text{tr}(x_1) \text{tr}(x_3), \text{tr}(x_2) \text{tr}(x_3). \end{aligned}$$

The SDP (6.2) minimizes over 31×31 positive semidefinite matrices subject to 881 linear equations in their entries. By solving it we get $\check{a}_{\min,2} = -0.0467$.

In the next step we have $d = 3$, and $\mathbf{M}_3^r(L)$ is a 108×108 matrix with 11270 linear relations. Now the solution of (6.2) is $\check{a}_{\min,3} = -0.0312$, which up to floating point precision agrees with $-\frac{1}{32}$. Since $\check{a}_{\min,3}$ is a lower bound for the solution of (6.1) and is attained by the 2×2 projections above, we conclude that $-\frac{1}{32}$ is the solution of (6.1).

Similar optimization problem can be used for detecting quantum entanglement [HHH01, BCS20]. Namely, trace polynomials are in correspondence with invariant operators [Hub21], and pure trace polynomials, positive subject to certain constraints, then relate to invariant operators positive on separable Werner states [Wer89, EW01]. Pure trace polynomial optimization can be thus used to efficiently produce entanglement witnesses for Werner states, which is a work in preparation.

6.2. Polynomial Bell inequalities. In this section we connect trace polynomial optimization to violations of nonlinear Bell inequalities, outline a few examples and prove the optimal bound on maximal violation of the covariance Bell inequality considered in [PHBB17], see Example 6.2.2 below.

As a prelude, the classical Bell inequality states that

$$(6.3) \quad \psi^*(A_1 \otimes B_1 + A_1 \otimes B_2 + A_2 \otimes B_1 - A_2 \otimes B_2)\psi$$

is at most 2 for all separable states $\psi \in \mathbb{C}^k \otimes \mathbb{C}^k$ and $A_j, B_j \in \mathbb{M}_k$ satisfying $A_j^* = A_j, A_j^2 = I, B_j^* = B_j, B_j^2 = I$. Tsirelson's bound implies that (6.3) is at most $2\sqrt{2}$ when arbitrary states are allowed. Moreover, the maximal value $2\sqrt{2}$ is attained when $k = 2$ and $\psi = \frac{1}{\sqrt{2}}(e_1 \otimes e_1 + e_2 \otimes e_2)$. In general, if ψ_k is the generalized Bell state,

$$\psi_k = \frac{1}{\sqrt{k}} \sum_{j=1}^k e_j \otimes e_j \in \mathbb{R}^k \otimes \mathbb{R}^k,$$

which is a maximally entangled bipartite state on $\mathbb{C}^k \otimes \mathbb{C}^k$, unique up to bipartite unitary equivalence, then

$$(6.4) \quad \psi_k^*(X \otimes Y)\psi_k = \text{tr}(XY)$$

for all $X, Y \in \mathbb{S}_k$. Therefore Tsirelson's bound for (6.3) on maximally entangled states can be recovered as a pure trace polynomial optimization problem

$$\sup \text{tr}(x_1y_1) + \text{tr}(x_1y_2) + \text{tr}(x_2y_1) - \text{tr}(x_2y_2) \quad \text{s.t.} \quad x_j^2 = 1, y_j^2 = 1.$$

Note that for finding the maximal violation for arbitrary $k \in \mathbb{N}$, there is no loss of generality if only symmetric matrices are considered instead of hermitian ones, since

$$\psi_k^*(Z \otimes W)\psi_k = \text{tr}(Z\overline{W}) = \text{tr} \left(\begin{pmatrix} \frac{Z+\overline{Z}}{2} & \frac{\overline{Z}-Z}{2i} \\ \frac{Z-\overline{Z}}{2i} & \frac{Z+\overline{Z}}{2} \end{pmatrix} \begin{pmatrix} \frac{W+\overline{W}}{2} & \frac{W-\overline{W}}{2i} \\ \frac{\overline{W}-W}{2i} & \frac{W+\overline{W}}{2} \end{pmatrix} \right)$$

for all hermitian $k \times k$ matrices Z, W .

Upper bounds on quantum violations of linear Bell inequalities can be found using the NPA hierarchy [NPA08] for eigenvalue optimization of noncommutative polynomials; for example, one can get Tsirelson's bound on violations of (6.3) by eigenvalue-optimizing $a_1b_1 + a_1b_2 + a_2b_1 - a_2b_2$ subject to $a_j^2 = b_j^2 = 1$ and $[a_i, b_j] = 0$.

On the other hand, bilocal models [BRGP12, Cha16], covariance of quantum correlations [PHBB17] and detection of partial separability [Uff02] lead to more general *polynomial* Bell inequalities. While linear Bell inequalities are linear in expectation values of (products of) observables, polynomial Bell inequalities contain multivariate polynomials in expectation values of (products of) observables. For this reason, noncommutative polynomial optimization is not suitable for studying violations of nonlinear Bell inequalities. In contrast, trace polynomial optimization gives upper bounds on violations of polynomial Bell inequalities, at least for certain families of states, e.g. the maximally entangled bipartite states via (6.4). We demonstrate this with the following examples.

6.2.1. *Example.* Consider a simple quadratic Bell inequality

$$(6.5) \quad (\psi^*(A_1 \otimes B_2 + A_2 \otimes B_1)\psi)^2 + (\psi^*(A_2 \otimes B_1 - A_2 \otimes B_2)\psi)^2 \leq 4$$

given in [Uff02], where it is shown that (6.5) holds for all separable states ψ , and for all 2-dimensional states (i.e. all states when $k = 2$). In [NKI02], (6.5) is shown to hold for arbitrary states, meaning it admits no quantum violations. An alternative automatized proof of (6.5) for maximally entangled states of arbitrary dimension can be obtained by solving the optimization problem

$$(6.6) \quad \sup (\text{tr}(x_1y_2 + x_2y_1))^2 + (\text{tr}(x_1y_1 - x_2y_2))^2 \quad \text{s.t.} \quad x_j^2 = 1, y_j^2 = 1 \text{ for } j = 1, 2$$

Let $S = \{\pm(1 - x_j^2), \pm(1 - y_j^2) : j = 1, 2\}$. The relaxation of (6.6) with $d = 2$ as in Section 5.3,

$$(6.7) \quad \inf \mu \quad \text{s.t.} \quad \mu - (\text{tr}(x_1y_2 + x_2y_1))^2 - (\text{tr}(x_1y_1 - x_2y_2))^2 \in \mathcal{M}(S(1))_d$$

outputs 4, which coincides with the classical value in (6.5). The concrete implementation of (6.7) encodes the relations $x_j^2 = y_j^2 = 1$ directly in the SDP using reduced words analogously as in the toy Example 6.1 (where projections were considered).

6.2.2. *Example.* Another class of polynomial Bell inequalities arises from covariances of quantum correlations. Let

$$\text{cov}_\psi(X, Y) = \psi^*(X \otimes Y)\psi - \psi^*(X \otimes I)\psi \cdot \psi^*(I \otimes Y)\psi$$

In [PHBB17] it is shown that while

$$(6.8) \quad \begin{aligned} & \text{cov}_\psi(A_1, B_1) + \text{cov}_\psi(A_1, B_2) + \text{cov}_\psi(A_1, B_3) \\ & + \text{cov}_\psi(A_2, B_1) + \text{cov}_\psi(A_2, B_2) - \text{cov}_\psi(A_2, B_3) \\ & + \text{cov}_\psi(A_3, B_1) - \text{cov}_\psi(A_3, B_2) \end{aligned}$$

is at most $\frac{9}{2}$ for separable states ψ , it attains the value 5 with the Bell state ψ_2 . The authors also performed extensive numerical search within entangled states for local dimensions $k \leq 5$, but no higher value of (6.8) was found. They leave it as an open question whether higher dimensional entangled states could lead to larger violations [PHBB17, Appendix D.1(b)].

Let

$$\begin{aligned} a = & \text{tr}(x_1 y_1) - \text{tr}(x_1) \text{tr}(y_1) + \text{tr}(x_1 y_2) - \text{tr}(x_1) \text{tr}(y_2) + \text{tr}(x_1 y_3) - \text{tr}(x_1) \text{tr}(y_3) \\ & + \text{tr}(x_2 y_1) - \text{tr}(x_2) \text{tr}(y_1) + \text{tr}(x_2 y_2) - \text{tr}(x_2) \text{tr}(y_2) - \text{tr}(x_2 y_3) + \text{tr}(x_2) \text{tr}(y_3) \\ & + \text{tr}(x_3 y_1) - \text{tr}(x_3) \text{tr}(y_1) - \text{tr}(x_3 y_2) + \text{tr}(x_3) \text{tr}(y_2). \end{aligned}$$

The relaxation of

$$(6.9) \quad \sup a \quad \text{s.t.} \quad x_j^2 = 1, y_j^2 = 1 \quad \text{for } j = 1, 2, 3$$

with $d = 2$ returns 5. Therefore the value of (6.8) is at most 5 for every maximally entangled state, regardless of the local dimension k .

6.2.3. *Example.* A family of eight quadratic Bell inequalities (arising from linear ones via elimination) corresponding to a bilocal model for three parties A, B, C is given in [Cha16]:

$$(6.10) \quad -\frac{1}{8}(J_1 \pm J_2)^2 - (\pm J_1 \pm J_2 + 2) \leq 0,$$

where

$$\begin{aligned} J_1 &= \sum_{i,j=1}^2 \psi'^*(A_i \otimes B'_1)\psi' \cdot \psi''^*(B''_1 \otimes C_j)\psi'', \\ J_2 &= \sum_{i,j=1}^2 (-1)^{i+j} \psi'^*(A_i \otimes B'_2)\psi' \cdot \psi''^*(B''_2 \otimes C_j)\psi'', \end{aligned}$$

and A_i, B'_j, B''_j, C_j are projections. Inequalities (6.10) are valid for every pair of separable states ψ', ψ'' , and are equivalent to $\sqrt{|J_1|} + \sqrt{|J_2|} \leq 1$ as derived in [BRGP12]. An upper bound on violations of (6.10) for maximally entangled shared states ψ', ψ'' is given by

$$(6.11) \quad \sup a \quad \text{s.t.} \quad x_j^2 = x_j, y_j'^2 = y_j', y_j''^2 = y_j'', z_j^2 = z_j \quad \text{for } j = 1, 2$$

where

$$a = -\frac{1}{8} \left(\sum_{i,j} \operatorname{tr}(x_i y'_1) \operatorname{tr}(y''_1 z_j) \pm \sum_{i,j} (-1)^{i+j} \operatorname{tr}(x_i y'_2) \operatorname{tr}(y''_2 z_j) \right)^2 \\ \pm \sum_{i,j} \operatorname{tr}(x_i y'_1) \operatorname{tr}(y''_1 z_j) \pm \sum_{i,j} (-1)^{i+j} \operatorname{tr}(x_i y'_2) \operatorname{tr}(y''_2 z_j) - 2.$$

While (6.11) fits in the trace polynomial optimization scheme presented in this paper, SDPs arising from (6.11) are very large because a is a pure trace polynomial of degree 8 in 8 variables. For computing upper bounds on (6.11) to become viable, the sizes of SDPs will need to be reduced using sparsity and symmetry techniques, which we plan to develop later.

7. CONCLUSION AND PERSPECTIVES

We have derived several novel Positivstellensätze for trace polynomials positive on tracial semialgebraic sets. Our tracial analog of Putinar's Positivstellensatz yields a converging hierarchy of semidefinite relaxations for optimizing pure trace polynomials under pure trace polynomial inequality constraints. We also provide an algorithm to extract minimizers of such problems, thanks to a finite-dimensional Gelfand-Naimark-Segal construction.

A topic of future research is to derive a hierarchy of primal-dual SDP programs converging to the minimal eigenvalue of a trace polynomial under trace polynomial inequality constraints. A short-term research investigation track is to rely on this hierarchy to tackle trace polynomial problems arising from quantum information theory. Sharing the same computational drawbacks as the classical Lasserre's hierarchy, our tracial framework will be limited to optimization problems involving a modest number of variables. To overcome this scalability issue, we intend to focus on exploiting structural properties of the input data. One possibility is to extend the framework from [KMP19] to optimization problems involving sparse trace polynomials or the one from [RTAL13] to problems involving symmetries.

Acknowledgments. The authors thank anonymous referees for their valuable comments and suggestions, which greatly improved presentation of the paper and demonstration of the main results.

A. ALTERNATIVE PROOF OF THEOREM 4.4

Proof of (i)⇒(iii). Assume $a + \varepsilon \notin \mathcal{M}^{\text{cyc}}$ for some $\varepsilon > 0$. Let $U = \{p \in \operatorname{Sym} \mathbb{T} \mid \operatorname{tr}(p) = 0\}$. Then $\mathcal{M}^{\text{cyc}} + U$ is a convex cone in $\operatorname{Sym} \mathbb{T}$. Since a is a pure trace polynomial, we have $a + \varepsilon \notin \mathcal{M}^{\text{cyc}} + U$. Since \mathcal{M}^{cyc} is archimedean, for every $p \in \operatorname{Sym} \mathbb{T}$ there exists $\delta > 0$ such that $1 \pm \delta p \in \mathcal{M}^{\text{cyc}}$, which in terms of [Bar02, Definition III.1.6] means that 1 is an algebraic interior point of the cone $\mathcal{M}^{\text{cyc}} + U$ in $\operatorname{Sym} \mathbb{T}$. By the Eidelheit-Kakutani separation theorem [Bar02, Corollary III.1.7] there is a nonzero \mathbb{R} -linear functional $L_0 : \operatorname{Sym} \mathbb{T} \rightarrow \mathbb{R}$ satisfying $L_0(\mathcal{M}^{\text{cyc}} + U) \subseteq \mathbb{R}_{\geq 0}$ and $L_0(a + \varepsilon) \leq 0$. In particular, $L_0(U) = \{0\}$. Moreover, $L_0(1) > 0$ because

\mathcal{M}^{cyc} is archimedean, so after rescaling we can assume $L_0(1) = 1$. Let $L : \mathbb{T} \rightarrow \mathbb{R}$ be the symmetric extension of L_0 , i.e., $L(p) = \frac{1}{2}L_0(p + p^*)$ for $p \in \mathbb{T}$. Note that

$$(A.1) \quad L(p) = L(\text{tr}(p))$$

for all $p \in \mathbb{T}$, and in particular $L(pq) = L(qp)$ for all $p, q \in \mathbb{T}$.

Now consider the set \mathcal{C} of all symmetric linear functionals $L' : \mathbb{T} \rightarrow \mathbb{R}$ satisfying $L'(\mathcal{M}^{\text{cyc}} + U) \subseteq \mathbb{R}_{\geq 0}$ and $L'(1) = 1$. This set is nonempty because $L \in \mathcal{C}$. Endow \mathbb{T} with the norm

$$\|p\| = \max \{ \|p(\underline{X})\| \mid n \in \mathbb{N}, \underline{X} \in \mathbb{S}_n^k, \|X_j\| \leq 1 \}.$$

This is indeed a norm because no nonzero trace polynomial vanishes on matrices of all finite sizes. By the Banach-Alaoglu theorem [Bar02, Theorem III.2.9], the convex set \mathcal{C} is weak*-compact. Thus by the Krein-Milman theorem [Bar02, Theorem III.4.1] we may assume that our separating functional L is an extreme point of \mathcal{C} .

On \mathbb{T} we define a semi-scalar product $\langle p, q \rangle = L(pq^*)$. By the Cauchy-Schwarz inequality for semi-scalar products,

$$\mathcal{N} = \{q \in \mathbb{T} \mid L(qq^*) = 0\}$$

is a linear subspace of \mathbb{T} . Let $p, q \in \mathbb{T}$. Since \mathcal{M}^{cyc} is archimedean, there exists $\delta > 0$ such that $1 - \delta pp^* \in \mathcal{M}^{\text{cyc}}$ and therefore

$$(A.2) \quad 0 \leq L(q(1 - \delta pp^*)q^*) = L(qq^*) - \delta L(qpp^*q^*) \leq L(qq^*).$$

In particular, $q \in \mathcal{N}$ implies $qp \in \mathcal{N}$, so \mathcal{N} is a left ideal. Furthermore, $L(\mathcal{N}) = \{0\}$: if $L(qq^*) = 0$, then for every $\delta > 0$,

$$0 \leq L((\delta \pm q)(\delta \pm q)^*) = \delta(\delta \pm 2L(q))$$

and hence $L(q) = 0$. Let $\bar{p} = p + \mathcal{N}$ denote the residue class of $p \in \mathbb{T}$ in \mathbb{T}/\mathcal{N} . Because \mathcal{N} is a left ideal, we can define linear maps

$$\chi_p : \mathbb{T}/\mathcal{N} \rightarrow \mathbb{T}/\mathcal{N}, \quad \bar{q} \mapsto \overline{pq}$$

for $p \in \mathbb{T}$, which are bounded by (A.2).

Now

$$(A.3) \quad \langle \bar{p}, \bar{q} \rangle = L(pq^*)$$

is a scalar product on \mathbb{T}/\mathcal{N} , and we let H denote the completion of \mathbb{T}/\mathcal{N} with respect to this scalar product. Each χ_p extends to a bounded operator $\hat{\chi}_p$ on H , and the map

$$(A.4) \quad \pi : \mathbb{T} \rightarrow \mathcal{B}(H), \quad p \mapsto \hat{\chi}_p$$

is clearly a \star -representation with $\ker \pi = \mathcal{N}$. Let \mathcal{F} be the closure of $\pi(\mathbb{T})$ in $\mathcal{B}(H)$ with respect to the weak operator topology. The map

$$\tau : \pi(\mathbb{T}) \rightarrow \mathbb{R}, \quad \hat{\chi}_p \mapsto L(p)$$

is a faithful tracial state on $\pi(\mathbb{T})$ by $L(\mathcal{N}) = \{0\}$ and (A.1). Since

$$\tau(\hat{\chi}_p) = \langle \bar{p}, \bar{1} \rangle,$$

τ extends uniquely to a faithful normal tracial state on \mathcal{F} .

Next we claim that $\pi(\mathbb{T}) = \mathbb{R}$. Observe that $\bar{1} \in H$ is a cyclic vector for π by construction and $L(p) = \langle \pi(p)\bar{1}, \bar{1} \rangle$. Suppose $\pi(\mathbb{T}) \neq \mathbb{R}$. If \mathcal{E} denotes the weak closure of $\pi(\mathbb{T})$ in \mathcal{F} , then \mathcal{E} is a central von Neumann subalgebra of \mathcal{F} ; since $\mathcal{E} \neq \mathbb{R}$ and all the elements of \mathcal{E} are self-adjoint, there is a nontrivial projection $P \in \mathcal{E}$. Since $\bar{1}$ is cyclic for π , we have $P\bar{1} \neq 0$ and $(1 - P)\bar{1} \neq 0$. Hence we can define linear functionals L_i on \mathbb{T} by

$$L_1(p) = \frac{\langle \pi(p)P\bar{1}, P\bar{1} \rangle}{\|P\bar{1}\|^2} \quad \text{and} \quad L_2(p) = \frac{\langle \pi(p)(1 - P)\bar{1}, (1 - P)\bar{1} \rangle}{\|(1 - P)\bar{1}\|^2}$$

for all $p \in \mathbb{T}$. One easily checks that L is a convex combination of L_1 and L_2 , $L_i(1) = 1$ and $L_i(\mathcal{M}^{\text{cyc}}) = \mathbb{R}_{\geq 0}$. Furthermore, since P is a weak limit of $\{\pi(s_n)\}$ for some $s_n \in \mathbb{T}$ and

$$\langle \pi(p - \text{tr}(p))\pi(s_n)\bar{1}, \bar{1} \rangle = L(s_n(p - \text{tr}(p))) = L(s_np - \text{tr}(s_np)) = 0,$$

we also have

$$\langle \pi(p - \text{tr}(p))P\bar{1}, P\bar{1} \rangle = \langle \pi(p - \text{tr}(p))P\bar{1}, \bar{1} \rangle = 0$$

so $L_i(U) = \{0\}$. Therefore $L_i \in \mathcal{C}$, so $L = L_1 = L_2$ by the extreme property of L . Then for $\lambda = \|P\bar{1}\|^2$,

$$\langle \pi(p)\bar{1}, \lambda\bar{1} \rangle = \lambda \langle \pi(p)\bar{1}, \bar{1} \rangle = \langle \pi(p)P\bar{1}, P\bar{1} \rangle = \langle P\pi(p)\bar{1}, P\bar{1} \rangle = \langle \pi(p)\bar{1}, P\bar{1} \rangle$$

for all $p \in \mathbb{T}$. Therefore $P\bar{1} = \lambda\bar{1}$ since $\bar{1}$ is a cyclic vector for π . So $\lambda \in \{0, 1\}$ since P is a projection, a contradiction.

Let $\underline{X} := (\hat{\chi}_{x_1}, \dots, \hat{\chi}_{x_n})$. This is a tuple of self-adjoint operators in \mathcal{F} , and $\pi(\mathbb{T}) = \mathbb{R}$ implies $p(\underline{X}) = \hat{\chi}_p$ for all $p \in \mathbb{T}$. Therefore $\underline{X} \in \mathcal{D}_{\mathcal{M}^{\text{cyc}}}^{\mathcal{F}, \tau}$ by (A.3) and $L(\mathcal{M}^{\text{cyc}}) \subseteq \mathbb{R}_{\geq 0}$. Finally $a(\underline{X}) = \tau(\hat{\chi}_a) = L(a) < 0$. \square

REFERENCES

- [AL12] Miguel F. Anjos and Jean B. Lasserre, editors. *Handbook on semidefinite, conic and polynomial optimization*, volume 166 of *International Series in Operations Research & Management Science*. Springer, New York, 2012. 2, 17
- [ARU97] Shavkat Ayupov, Abdugafur Rakhimov, and Shukhrat Usmanov. *Jordan, real and Lie structures in operator algebras*, volume 418 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1997. 6
- [Bar02] Alexander Barvinok. *A course in convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002. 26, 27
- [BCKP13] Sabine Burgdorf, Kristijan Cafuta, Igor Klep, and Janez Povh. The tracial moment problem and trace-optimization of polynomials. *Math. Program.*, 137(1-2, Ser. A):557–578, 2013. 2, 17
- [BCS20] Ivan Bardet, Benoît Collins, and Gunjan Sapra. Characterization of equivariant maps and application to entanglement detection. *Ann. Henri Poincaré*, 21(10):3385–3406, 2020. 23
- [Bel64] John S Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964. 2
- [BKP16] Sabine Burgdorf, Igor Klep, and Janez Povh. *Optimization of polynomials in non-commuting variables*. SpringerBriefs in Mathematics. Springer, [Cham], 2016. 2, 6, 12, 14, 17, 18
- [BMV75] Daniel Bessis, Pierre Moussa, and Matteo Villani. Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics. *J. Math. Phys.*, 16(11):2318–2325, 1975. 1

- [BRGP12] Cyril Branciard, Denis Rosset, Nicolas Gisin, and Stefano Pironio. Bilocal versus non-bilocal correlations in entanglement-swapping experiments. *Phys. Rev. A*, 85:032119, Mar 2012. 24, 25
- [CF98] Raúl E. Curto and Lawrence A. Fialkow. Flat extensions of positive moment matrices: recursively generated relations. *Mem. Amer. Math. Soc.*, 136(648):x+56, 1998. 17
- [Cha16] Rafael Chaves. Polynomial Bell inequalities. *Phys. Rev. Lett.*, 116(1):010402, 6, 2016. 24, 25
- [CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. rev. lett.*, 23(15):880, 1969. 2
- [CKP11] Kristijan Cafuta, Igor Klep, and Janez Povh. NCSOSTools: a computer algebra system for symbolic and numerical computation with noncommutative polynomials. *Optim. Methods Softw.*, 26(3):363–380, 2011. 2
- [CKP12] Kristijan Cafuta, Igor Klep, and Janez Povh. Constrained polynomial optimization problems with noncommuting variables. *SIAM J. Optim.*, 22(2):363–383, 2012. 2, 16
- [DLTW08] Andrew C. Doherty, Yeong-Cherng Liang, Ben Toner, and Stephanie Wehner. The quantum moment problem and bounds on entangled multi-prover games. In *2008 23rd Annual IEEE Conference on Computational Complexity*, pages 199–210. IEEE, 2008. 2
- [dOHMP09] Mauricio C. de Oliveira, J. William Helton, Scott A. McCullough, and Mihai Putinar. Engineering systems and free semi-algebraic geometry. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 17–61. Springer, New York, 2009. 1
- [Dyk94] Kenneth J. Dykema. Factoriality and Connes’ invariant $T(\mathcal{M})$ for free products of von Neumann algebras. *J. Reine Angew. Math.*, 450:159–180, 1994. 7
- [EW01] Tilo Eggeling and Reinhard F. Werner. Separability properties of tripartite states with $U \otimes U \otimes U$ symmetry. *Phys. Rev. A (3)*, 63(4):042111, 15, 2001. 23
- [FN14] Motohisa Fukuda and Ion Nechita. Asymptotically well-behaved input states do not violate additivity for conjugate pairs of random quantum channels. *Comm. Math. Phys.*, 328(3):995–1021, 2014. 3
- [GdLL18] Sander Gribling, David de Laat, and Monique Laurent. Bounds on entanglement dimensions and quantum graph parameters via noncommutative polynomial optimization. *Math. Program.*, 170(1, Ser. B):5–42, 2018. 1
- [GdLL19] Sander Gribling, David de Laat, and Monique Laurent. Lower bounds on matrix factorization ranks via noncommutative polynomial optimization. *Found. Comput. Math.*, to appear 2019. 1
- [Had01] Don Hadwin. A noncommutative moment problem. *Proc. Amer. Math. Soc.*, 129(6):1785–1791, 2001. 7, 8
- [Hel02] J. William Helton. “Positive” noncommutative polynomials are sums of squares. *Ann. of Math. (2)*, 156(2):675–694, 2002. 2
- [HHH01] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of n -particle mixed states: necessary and sufficient conditions in terms of linear maps. *Physics Letters A*, 283(1):1–7, 2001. 23
- [HKM11] J. William Helton, Igor Klep, and Scott McCullough. Proper analytic free maps. *J. Funct. Anal.*, 260(5):1476–1490, 2011. 6
- [HKT17] Fumio Hiai, Robert König, and Marco Tomamichel. Generalized log-majorization and multivariate trace inequalities. *Ann. Henri Poincaré*, 18(7):2499–2521, 2017. 1
- [HM04] J. William Helton and Scott A. McCullough. A Positivstellensatz for noncommutative polynomials. *Trans. Amer. Math. Soc.*, 356(9):3721–3737, 2004. 2, 4, 6, 9, 11, 12
- [Hub21] Felix Huber. Positive maps and trace polynomials from the symmetric group. *J. Math. Phys.*, 62(2):022203, 2021. 1, 3, 23

- [JNV⁺20] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. $\text{MIP}^* = \text{RE}$. *arXiv preprint arXiv:2001.04383*, 2020. 5
- [KMP19] Igor Klep, Victor Magron, and Janez Povh. Sparse noncommutative polynomial optimization. *arXiv preprint arXiv:1909.00569*, 2019. 26
- [KPV18] Igor Klep, Janez Povh, and Jurij Volčič. Minimizer extraction in polynomial optimization is robust. *SIAM J. Optim.*, 28(4):3177–3207, 2018. 20
- [KS08] Igor Klep and Markus Schweighofer. Connes’ embedding conjecture and sums of Hermitian squares. *Adv. Math.*, 217(4):1816–1837, 2008. 5
- [KŠ17] Igor Klep and Špela Špenko. Free function theory through matrix invariants. *Canad. J. Math.*, 69(2):408–433, 2017. 5
- [KŠV18] Igor Klep, Špela Špenko, and Jurij Volčič. Positive trace polynomials and the universal Procesi–Schacher conjecture. *Proc. Lond. Math. Soc.*, 117(6):1101–1134, 2018. 3, 5, 8
- [Lam13] Tsit-Yuen Lam. *A first course in noncommutative rings*, volume 131. Springer Science & Business Media, 2013. 18
- [Las01] Jean-Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2000/01. 2
- [Lau09] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 157–270. Springer, New York, 2009. 2
- [Lax58] Peter D. Lax. Differential equations, difference equations and matrix theory. *Comm. Pure Appl. Math.*, 11:175–194, 1958. 1
- [LLR08] Jean Bernard Lasserre, Monique Laurent, and Philipp Rostalski. Semidefinite characterization and computation of zero-dimensional real radical ideals. *Foundations of Computational Mathematics*, 8(5):607–647, 2008. 18
- [LS04] Elliott H. Lieb and Robert Seiringer. Equivalent forms of the Bessis-Moussa-Villani conjecture. *J. Statist. Phys.*, 115(1-2):185–190, 2004. 1
- [Mar08] Murray Marshall. *Positive polynomials and sums of squares*, volume 146 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2008. 4, 7, 12
- [McC01] Scott McCullough. Factorization of operator-valued polynomials in several noncommuting variables. *Linear Algebra Appl.*, 326(1-3):193–203, 2001. 2, 17
- [MKKK10] Kazuo Murota, Yoshihiro Kanno, Masakazu Kojima, and Sadayoshi Kojima. A numerical algorithm for block-diagonal decomposition of matrix $*$ -algebras with application to semidefinite programming. *Jpn. J. Ind. Appl. Math.*, 27(1):125–160, 2010. 19
- [NKI02] Koji Nagata, Masato Koashi, and Nobuyuki Imoto. Configuration of separability and tests for multipartite entanglement in Bell-type experiments. *Phys. Rev. Lett.*, 89(26):260401, 4, 2002. 24
- [NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New J. Phys.*, 10(7):073013, 2008. 1, 2, 24
- [NT14] Tim Netzer and Andreas Thom. Hyperbolic polynomials and generalized Clifford algebras. *Discrete Comput. Geom.*, 51(4):802–814, 2014. 1
- [PHBB17] Victor Pozsgay, Flavien Hirsch, Cyril Branciard, and Nicolas Brunner. Covariance Bell inequalities. *Phys. Rev. A*, 96(6):062128, 13, 2017. 23, 24, 25
- [PKRR⁺19] Alejandro Pozas-Kerstjens, Rafael Rabelo, Lukasz Rudnicki, Rafael Chaves, Daniel Cavalcanti, Miguel Navascués, and Antonio Acín. Bounding the sets of classical and quantum correlations in networks. *Phys. Rev. Lett.*, 123(14):140503, 6, 2019. 1, 3
- [PNA10] Stefano Pironio, Miguel Navascués, and Antonio Acín. Convergent relaxations of polynomial optimization problems with noncommuting variables. *SIAM J. Optim.*, 20(5):2157–2180, 2010. 2, 17

- [Pro76] Claudio Procesi. The invariant theory of $n \times n$ matrices. *Adv. Math.*, 19(3):306–381, 1976. 5, 16
- [PS98] George Pólya and Gabor Szegő. *Problems and theorems in analysis. II*. Classics in Mathematics. Springer-Verlag, Berlin, 1998. Theory of functions, zeros, polynomials, determinants, number theory, geometry, Translated from the German by C. E. Billigheimer, Reprint of the 1976 English translation. 12
- [Put93] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.*, 42(3):969–984, 1993. 2, 4
- [PV09] Károly F. Pál and Tamás Vértesi. Quantum bounds on Bell inequalities. *Phys. Rev. A* (3), 79(2):022120, 12, 2009. 2
- [RTAL13] Cordian Riener, Thorsten Theobald, Lina Jansson Andrén, and Jean-Bernard Lasserre. Exploiting symmetries in SDP-relaxations for polynomial optimization. *Math. Oper. Res.*, 38(1):122–141, 2013. 26
- [SBT17] David Sutter, Mario Berta, and Marco Tomamichel. Multivariate trace inequalities. *Comm. Math. Phys.*, 352(1):37–58, 2017. 1
- [Sch91] Konrad Schmüdgen. The K-moment Problem for Compact Semi-algebraic Sets. *Mathematische Annalen*, 289(2):203–206, 1991. 10
- [SIG98] Robert E. Skelton, Tetsuya Iwasaki, and Karolos M. Grigoriadis. *A unified algebraic approach to linear control design*. The Taylor & Francis Systems and Control Book Series. Taylor & Francis, Ltd., London, 1998. 1
- [Sta13] Herbert R. Stahl. Proof of the BMV conjecture. *Acta Math.*, 211(2):255–290, 2013. 1
- [Tak02] Masamichi Takesaki. *Theory of operator algebras. I*, volume 124 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2002. Reprint of the first (1979) edition, Operator Algebras and Non-commutative Geometry, 5. 6, 7, 11
- [Uff02] Jos Uffink. Quadratic Bell inequalities as tests for multipartite entanglement. *Phys. Rev. Lett.*, 88(23):230406, 4, 2002. 24
- [Vid59] Ivan Vidav. On some *-regular rings. *Acad. Serbe Sci. Publ. Inst. Math.*, 13:73–80, 1959. 10
- [Wer89] Reinhard F. Werner. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, Oct 1989. 23
- [WSV12] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenbergh. *Handbook of semidefinite programming: theory, algorithms, and applications*, volume 27. Springer Science & Business Media, 2012. 18

IGOR KLEP: FACULTY OF MATHEMATICS AND PHYSICS, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF LJUBLJANA, SLOVENIA

E-mail address: igor.klep@fmf.uni-lj.si

VICTOR MAGRON: LAAS-CNRS & INSTITUTE OF MATHEMATICS FROM TOULOUSE, FRANCE

E-mail address: vmagron@laas.fr

JURIJ VOLČIČ: DEPARTMENT OF MATHEMATICS, TEXAS A&M UNIVERSITY, TEXAS

E-mail address: volcic@math.tamu.edu

NOT FOR PUBLICATION

CONTENTS

1. Introduction	1
Contributions	3
2. Notation and basic definitions	5
2.1. Noncommutative polynomials and trace polynomials	5
2.2. Tracial semialgebraic sets and von Neumann algebras	5
3. Non-cyclic Positivstellensatz for pure trace polynomials	7
4. Cyclic Positivstellensatz for trace polynomials	9
5. SDP hierarchy for trace optimization	13
5.1. SDP hierarchy for pure trace polynomial optimization	14
5.2. Finite-dimensional GNS representations and minimizer extraction	17
5.3. SDP hierarchy for trace polynomial optimization	20
6. Examples and applications	22
6.1. A toy example	22
6.2. Polynomial Bell inequalities	23
7. Conclusion and perspectives	26
Acknowledgments	26
Appendix A. Alternative proof of Theorem 4.4	26
References	28