STATE POLYNOMIALS: POSITIVITY, OPTIMIZATION AND NONLINEAR BELL INEQUALITIES

IGOR KLEP, VICTOR MAGRON, JURIJ VOLČIČ, AND JIE WANG

Abstract. This paper introduces state polynomials, i.e., polynomials in noncommuting variables and formal states of their products. A state analog of Artin's solution to Hilbert's 17th problem is proved showing that state polynomials, positive over all matrices and matricial states, are sums of squares with denominators. Somewhat surprisingly, it is also established that a Krivine-Stengle Positivstellensatz fails to hold in the state polynomial setting. Further, archimedean Positivstellensätze in the spirit of Putinar and Helton-McCullough are presented leading to a hierarchy of semidefinite relaxations converging monotonically to the optimum of a state polynomial subject to state constraints. This hierarchy can be seen as a state analog of the Lasserre hierarchy for optimization of polynomials, and the Navascués-Pironio-Acín scheme for optimization of noncommutative polynomials. The motivation behind this theory arises from the study of correlations in quantum networks. Determining the maximal quantum violation of a polynomial Bell inequality for an arbitrary network is reformulated as a state polynomial optimization problem. Several examples of quadratic Bell inequalities in the bipartite and the bilocal tripartite scenario are analyzed. To reduce the size of the constructed SDPs, sparsity, sign symmetry and conditional expectation of the observables' group structure are exploited. To obtain the above-mentioned results, techniques from noncommutative algebra, real algebraic geometry, operator theory, and convex optimization are employed.

2020 *Mathematics Subject Classification.* 13J30, 46L30, 90C22, 46N50, 81-08.

Key words and phrases. Noncommutative polynomial, state polynomial, Hilbert's 17th problem, Positivstellensatz, state optimization, semidefinite programming, network scenario, polynomial Bell inequality.

IK was supported by the Slovenian Research Agency grants J1-2453, N1-0217, J1-3004 and P1-0222.

VM was supported by the EPOQCS grant funded by the LabEx CIMI (ANR-11-LABX-0040), the FastQI grant funded by the Institut Quantique Occitan, the PHC Proteus grant 46195TA, the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions, grant agreement 813211 (POEMA), by the AI Interdisciplinary Institute ANITI funding, through the French "Investing for the Future PIA3" program under the Grant agreement n° ANR-19-PI3A-0004 as well as by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

JV was supported by the NSF grant DMS-1954709.

JW was supported by the NSFC grant 12201618.

1. Introduction

This paper introduces the class of *(noncommutative) state polynomials*, i.e., polynomials in noncommutative (nc) variables, such as matrices or operators, and formal states of their products. Such polynomials are naturally evaluated over finite or infinite-dimensional Hilbert spaces $\mathcal{H}$ by replacing each variable by a bounded operator on $\mathcal{H}$, and picking a state, i.e., a positive unital linear functional on the set of bounded operators $\mathcal{B}(\mathcal{H})$. The aim of the paper is to study positivity and optimization of state polynomials, and develop corresponding algebraic positivity certificates and associated algorithms. The main motivation for studying state polynomials arises from quantum information theory, in particular nonlinear Bell inequalities [Cha16, PHBB17] for correlations in quantum networks [Fri12, PKRR+19, TPKLR22]. Namely, it turns out that computing the maximum quantum violation of a polynomial Bell inequality in the standard Bell scenario corresponds to optimizing a state polynomial under nc (in)equality constraints; that is, constraints only involve nc variables and not the state. For more general quantum networks, polynomial Bell inequalities correspond to state polynomial optimization problems subject to both nc and state (in)equalities.

In the free nc context, i.e., in the absence of states, several representation results for positive polynomials (or Positivstellensätze) have been derived, allowing one to perform optimization. One of the central results from Helton and McCullough independently [Hel02, McC01] asserts that all positive semidefinite polynomials are *sums of hermitian squares* (SOHS). This in turn allows one to minimize the eigenvalue of an nc polynomial. One can also minimize the eigenvalue of an nc polynomial subject to a finite number of nc polynomial inequality constraints, i.e., over a basic nc semialgebraic set. More precisely, a non-decreasing sequence of lower bounds of the minimal eigenvalue can be obtained, each bound corresponding to the solution of a semidefinite program (SDP)[1]. Thanks to the Helton-McCullough representation theorem [HM04], the corresponding hierarchy of lower bounds converges to the minimal eigenvalue if the quadratic module generated by the polynomials describing the basic nc semialgebraic set is archimedean. This framework is the nc variant of the nowadays famous Lasserre's hierarchy [Las01] for *commutative* polynomial optimization, based on the representation by Putinar [Put93] of positive polynomials over basic closed semialgebraic sets. Hierarchies of semidefinite programs have been applied and generalized to different nc optimization problems [HM04, NPA08, PNA10, BCKP13]. In the seminal paper [NPA08], Navascués, Pironio and Acín (NPA) provide such a hierarchy to bound the maximal violation levels of linear Bell inequalities after casting the initial quantum information problem as an eigenvalue maximization problem; cf. [DLTW08]. Extensions to trace minimization of nc polynomials have been derived in [BCKP13]. More recently, several hierarchies have been derived in [GdLL19, GLS22] to provide lower bounds for various matrix factorization ranks. These hierarchies have been concretely implemented in the Matlab library `NCSOStools` [BKP16] and the Julia library `TSSOS` [MW23, Appendix B].

[1]That is, the optimum of a linear function subject to linear matrix inequality (LMI) constraints.

Recent efforts significantly extend these frameworks to the case of optimization problems involving *trace* polynomials, i.e., polynomials in nc variables and traces of their products. In [KŠV18], the first and thirds authors focused on trace polynomials being positive on semialgebraic sets of *fixed size* matrices, and derived several Positivstellensätze, including a Putinar-type Positivstellensatz stating that any positive trace polynomial admits a weighted SOHS decomposition without denominators. In [KMV22], the first, second and third authors generalized the above framework to the free setting, by providing a Putinar-type Positivstellensatz for trace polynomials which are positive on tracial semialgebraic sets, where the evaluations are performed on von Neumann algebras. This latter framework was applied in [HKMV22] to detect entanglement of Werner state witnesses in a dimension-free way. In the univariate case, a tracial analog of Artin's solution to Hilbert's 17th problem was provided in [KPV21], where it is proved that a positive semidefinite univariate trace polynomial is a quotient of sums of products of squares and traces of squares of trace polynomials. In the multivariate unconstrained setting, it is shown in [KSV22] that trace-positive nc polynomials can be "weakly" approximated by SOHS and commutators of regular nc rational functions.

From the point of view of quantum information, the trace polynomial optimization framework from [KMV22] allows us to obtain bounds on violation levels of nonlinear Bell inequalities corresponding to *maximally* entangled states. In this paper we rely on state polynomial optimization that is less restrictive, as it can provide violation bounds reached by (not necessary maximally) entangled states. From the point of view of operator theory, there is a correspondence between states on a Hilbert space $\mathcal{H}$ and trace-class operators on $\mathcal{H}$, but the reformulation of a state polynomial optimization problem into one with trace polynomials involves the *non-normalized trace*, in which case there is no dimension-independent theory of positivity, necessitating the introduction of this new class of objects, i.e., (nc) state polynomials.

Contributions and main results. A *state polynomial* in nc variables $x_1, \dots, x_n$ is a real polynomial in formal state symbols $\varsigma(w)$, where w is a word in $x_1, \dots, x_n$. More generally, an *nc state polynomial* is a polynomial in $x_1, \dots, x_n$ and formal states of their words. For example, $f = \varsigma(x_1x_2x_1) - \varsigma(x_1)\varsigma(x_1x_2)$ is a state polynomial, and $h = \varsigma(x_1^2)x_2x_1 + \varsigma(x_1)\varsigma(x_2x_1x_2)$ is an nc state polynomial. At a pair of bounded operators $\underline{X} = (X_1, X_2)$ on Hilbert space $\mathcal{H}$ and a state λ on $\mathcal{B}(\mathcal{H})$, they are evaluated as $f(\lambda; \underline{X}) = \lambda(X_1X_2X_1) - \lambda(X_1)\lambda(X_1X_2)$ and $h(\lambda; \underline{X}) = \lambda(X_1^2)X_2X_1 + \lambda(X_1)\lambda(X_2X_1X_2)I$.

State polynomials form a commutative algebra denoted $\mathscr{S}$, and nc state polynomials form a noncommutative algebra denoted $\boldsymbol{\mathscr{S}}$. There is a canonical involution $\star$ on $\boldsymbol{\mathscr{S}}$ that fixes $\mathscr{S} \cup \{x_1, \dots, x_n\}$ element-wise, and an $\mathscr{S}$-linear map $\varsigma : \boldsymbol{\mathscr{S}} \to \mathscr{S}$.

After establishing the algebraic framework for state polynomials in Section 2 and their function theoretic perspective in Section 3, we prove our first main result, the affirmative answer to a state polynomial analog of Hilbert's 17th problem from real algebraic geometry [Mar08, Sch09].

Theorem A (Theorem 4.3)**.** *Let f be a state polynomial. Then $f(\lambda;\underline{X}) \geq 0$ for all matricial states λ and tuples of symmetric matrices $\underline{X}$ if and only if f is a quotient of sums of products of elements of the form $\varsigma(hh^\star)$ for an nc state polynomial h.*

For example,

$$\varsigma(x_1^2)\varsigma(x_2^2) - \varsigma(x_1x_2)^2 = \frac{\varsigma\Big(\big(\varsigma(x_1^2)x_2 - \varsigma(x_1x_2)x_1\big)^2\Big)}{\varsigma(x_1)^2}$$

is an algebraic certificate for the Cauchy-Schwarz inequality, a sum of hermitian squares (SOHS) certificate of the form guaranteed for all global state polynomial inequalities by Theorem A. As a consequence of Theorem A, positivity of a state polynomial on all matrix tuples and matricial states implies positivity on all bounded operators and states.

After global positivity, we turn to constrained positivity of state polynomials in Section 5. We restrict ourselves to constraint sets $C \subseteq \mathscr{S}$ that are *balanced*: namely, C is closed under the involution $\star$, and the non-symmetric elements of C come in pairs with their negatives (to allow us to handle equality constraints). Let $\mathcal{H}$ be a separable real Hilbert space. Given a balanced set $C \subseteq \mathscr{S}$, let $\mathcal{D}_C^\infty$ be the set of all pairs $(\lambda;\underline{X})$ of a state λ on $\mathcal{B}(\mathcal{H})$ and $\underline{X} \in \mathcal{B}(\mathcal{H})^n$ such that $c(\lambda;\underline{X}) \succeq 0$ for all $c \in C$. We call $\mathcal{D}_C^\infty$ the *state semialgebraic set* constrained by C. While Theorem A gives an SOHS certificate for global positivity of a state polynomial (even when only matrix evaluations are considered), there is no comparable analog for positivity on arbitrary state semialgebraic sets. In Section 5.2 we show that the state versions of some of the classic (Krivine-Stengle and Schmüdgen) Positivstellensätze fail in general. Nevertheless, there is an analog of Putinar's archimedean Positivstellensatz [Put93]. We say that $C \subseteq \mathscr{S}$ is *algebraically bounded* if $N - x_1^2 - \cdots - x_n^2 = \sum_i p_i c_i p_i^\star$ for some $c_i \in C$ and nc polynomials p_i. Note that $\mathcal{D}_C^\infty$ for an algebraically bounded C is bounded in operator norm; conversely, if $\mathcal{D}_C^\infty$ is bounded, then one can make C algebraically bounded without changing the state semialgebraic set $\mathcal{D}_C^\infty$ by adding a single constraint. For algebraically-bounded constraint sets, we obtain the following Positivstellensatz.

Theorem B (Theorem 5.5)**.** *Let f be a state polynomial, and C a balanced algebraically bounded set of nc state polynomials. Then $f \geq 0$ on $\mathcal{D}_C^\infty$ if and only if for every $\varepsilon > 0$,*

$$f + \varepsilon = \sum_i \varsigma(h_i c_i h_i^\star)$$

for some nc state polynomials h_i and $c_i \in \{1\} \cup C$.

Theorem B is the cornerstone of the state optimization framework we develop in Section 6. For a state polynomial f and a balanced algebraically bounded set $C \subseteq \mathscr{S}$, Theorem B gives rise to an SDP hierarchy that produces a convergent increasing sequence with limit $\inf_{\mathcal{D}_C^\infty} f$ (Corollary 6.1). Under a mild condition on C, these SDPs satisfy strong duality (Proposition 6.7). Furthermore, under flatness and extremal assumptions, the dual SDPs and a variant of the Gelfand-Naimark-Segal construction allow us to extract a finite-dimensional minimizer for $\inf_{\mathcal{D}_C^\infty} f$, and thus obtain finite convergence of our hierarchy (Proposition 6.10). The complexity of the involved SDPs grows rather quickly; nevertheless, we can exploit correlative sparsity (Theorem 6.12) and sign symmetry (Theorem 6.13) patterns in f and C to reduce the SDP sizes considerably.

Finally, we apply our newly developed theory to quantum correlations in networks. Section 7 considers nonlinear Bell inequalities [Uff02, Cha16] in the standard Bell scenario, where two parties share an entanglement source. While a linear Bell inequality for quantum (commuting) correlations in such a scenario corresponds to eigenvalue optimization of an nc polynomial, a polynomial Bell inequality for quantum correlations corresponds to a state polynomial optimization problem. The form of the constraints arising from the quantum mechanical formalism allows for a further reduction of the size of the obtained SDPs in the hierarchy using a conditional expectation induced by the underlying group structure of binary observables (Proposition 7.1). Section 8 generalizes the aforementioned correspondence to correlation inequalities in general network scenarios [Fri12]. That is, several entanglement sources and sharing patterns are permitted. Following [PKRR+19, LGG21, RX22], quantum commuting models for a network can be characterized using state polynomial constraints. This allows us to apply our optimization results to analyze polynomial Bell inequalities for correlations in arbitrary networks.

Theorem C (Corollaries 6.1 and 8.1)**.** *The largest quantum violation of a polynomial Bell inequality for classical correlations in a network scenario is the limit of a convergent decreasing sequence produced by the Positivstellensatz-generated SDP hierarchy.*

Using the derived optimization tools, we establish novel largest quantum violations or their nontrivial upper bounds for various polynomial Bell inequalities in the bipartite scenario (Section 7.2) and in the bilocal tripartite scenario (Section 8.1) from the literature.

Acknowledgements. The authors thank Timotej Hrga for sharing his expertise and solving the large semidefinite program pertaining to one of the Bell inequality examples.

2. Preliminaries

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

2.1. **Noncommutative polynomials and state polynomials.** Let $\mathrm{S}_k(\mathbb{R})$ denote the space of all real symmetric matrices of order k. For a set A, we use $|A|$ to denote its cardinality. 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 $\underline{x} = (x_1, \dots, x_n)$. Let $|w|$ denote the length of $w \in \langle \underline{x} \rangle$. We denote by $\mathbb{R}\langle \underline{x} \rangle$ the set of real polynomials in noncommutative variables, abbreviated as *nc polynomials*. The free 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 variables $x_1, \dots, x_n$.

For $w \in \langle \underline{x} \rangle \setminus \{1\}$ let $\varsigma(w)$ be a symbol subject to the relation $\varsigma(w) = \varsigma(w^\star)$, and let

$$\mathscr{S} := \mathbb{R}\big[\varsigma(w) \colon w \in \langle \underline{x} \rangle \setminus \{1\}\big],$$

a commutative polynomial ring in infinitely many variables. An element in $\mathscr{S}$ of the form $\prod_{j=1}^m \varsigma(u_j)$ for $u_j \in \langle \underline{x} \rangle \setminus \{1\}$ is called an *$\mathscr{S}$-word*. The set of all $\mathscr{S}$-words is a vector space basis of $\mathscr{S}$. The *degree* of an $\mathscr{S}$-word $\prod_j \varsigma(u_j)$ equals $\sum_j |u_j|$. The vector of $\mathscr{S}$-words whose degrees are no greater than d is denoted by $W_d^{\mathscr{S}}$.

We also let $\boldsymbol{\mathscr{S}} := \mathscr{S} \otimes \mathbb{R}\langle \underline{x} \rangle$ be the free $\mathscr{S}$-algebra on $\underline{x}$. Elements of $\mathscr{S}$ are called *state polynomials*, and elements of $\boldsymbol{\mathscr{S}}$ are *nc state polynomials*. For example, $\varsigma(x_1x_2) - \varsigma(x_1)\varsigma(x_2) \in \mathscr{S}$ and $x_1x_2x_1 - \varsigma(x_2)x_1 + \varsigma(x_1x_2) - \varsigma(x_1)\varsigma(x_2) \in \boldsymbol{\mathscr{S}} = \mathscr{S}\langle x_1, x_2 \rangle$. An nc state polynomial of the form $\prod_{j=1}^m \varsigma(u_j)v$ for $u_j \in \langle \underline{x} \rangle \setminus \{1\}$ and $v \in \langle \underline{x} \rangle$ is called an *$\boldsymbol{\mathscr{S}}$-word*. The set of all $\boldsymbol{\mathscr{S}}$-words is a vector space basis of $\boldsymbol{\mathscr{S}}$. The *degree* of an $\boldsymbol{\mathscr{S}}$-word $\prod_j \varsigma(u_j)v$ equals $|v| + \sum_j |u_j|$. The *degree* of $f \in \boldsymbol{\mathscr{S}}$ is the maximal degree of $\boldsymbol{\mathscr{S}}$-words in the expansion of f.

The involution on $\boldsymbol{\mathscr{S}}$, denoted also by $\star$, fixes $\{x_1, \dots, x_n\} \cup \mathscr{S}$ point-wise, and reverses words from $\langle \underline{x} \rangle$. The set of all *symmetric elements* of $\boldsymbol{\mathscr{S}}$ is defined as $\operatorname{Sym} \boldsymbol{\mathscr{S}} := \{f \in \boldsymbol{\mathscr{S}} : f = f^\star\}$. We also consider the unital $\mathscr{S}$-linear $\star$-map $\varsigma : \boldsymbol{\mathscr{S}} \to \mathscr{S}$ uniquely determined by $w \mapsto \varsigma(w)$ for $w \in \langle \underline{x} \rangle \setminus \{1\}$.

2.2. **State semialgebraic sets.** Let $\mathcal{H}$ be a real Hilbert space, and let $\mathcal{S}(\mathcal{H})$ be the set of all *states* (positive unital $\star$-linear bounded functionals) on $\mathcal{B}(\mathcal{H})$. Each state arises from a positive-semidefinite trace-class operator $\rho \in \mathcal{B}(\mathcal{H})$ with trace 1, via $Y \mapsto \operatorname{tr}(\rho Y)$ (here, $\operatorname{tr} X = \sum_j \langle X e_j, e_j \rangle$ for a trace-class operator X, where $(e_j)_j$ is an arbitrary orthonormal basis of $\mathcal{H}$). Every unit vector $v \in \mathcal{H}$ determines a *vector state* $Y \mapsto \langle Yv, v \rangle$. Given an nc state polynomial $a \in \boldsymbol{\mathscr{S}}$, a state $\lambda \in \mathcal{S}(\mathcal{H})$, and a tuple $\underline{X} = (X_1, \dots, X_n)$ of self-adjoint operators $X_j = X_j^* \in \mathcal{B}(\mathcal{H})$, there is a natural evaluation

$$a(\lambda; \underline{X}) \in \mathcal{B}(\mathcal{H})$$

obtained by replacing w with $w(X) \in \mathcal{B}(\mathcal{H})$ and $\varsigma(w)$ with $\lambda(w(X)) \in \mathbb{R}$. Equivalently, each pair $(\lambda; \underline{X})$ gives rise to a $\star$-representation $\boldsymbol{\mathscr{S}} \to \mathcal{B}(\mathcal{H})$ that intertwines $\varsigma : \boldsymbol{\mathscr{S}} \to \mathscr{S}$ and $\lambda : \mathcal{B}(\mathcal{H}) \to \mathbb{R}$.

Throughout the text $\mathcal{H}$ is a separable infinite-dimensional Hilbert space (e.g. $\mathcal{H} = \ell^2$).

Definition 2.1. A set $C \subseteq \boldsymbol{\mathscr{S}}$ is *balanced* if $C^\star = C$ and $-(C \setminus \operatorname{Sym} \boldsymbol{\mathscr{S}}) \subseteq C$. Given a balanced C let

$$\mathcal{D}_C^\infty := \{(\lambda, \underline{X}) \in \mathcal{S}(\mathcal{H}) \times \mathcal{B}(\mathcal{H})^n : X_j = X_j^*,\ c(\lambda; \underline{X}) \succeq 0 \text{ for all } c \in C\} \tag{2.1}$$

and

$$\mathcal{D}_C := \bigcup_{k \in \mathbb{N}} \{(\lambda, \underline{X}) \in \mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n : c(\lambda; \underline{X}) \succeq 0 \text{ for all } c \in C\}. \tag{2.2}$$

Let $\vec{\mathcal{D}}_C^\infty$ and $\vec{\mathcal{D}}_C$ be the analogs of (2.1) and (2.2), respectively, where one restricts only to vector states.

Note that any subset of Sym $\mathscr{S}$ is an example of a balanced set. The motivation behind more general balanced sets is to allow for non-symmetric *equality* constraints, for example commutation relations. Indeed, every non-symmetric element c in a balanced set contributes inequalities $c(\lambda;\underline{X}) \succeq 0$ and $-c(\lambda;\underline{X}) \succeq 0$, and thus $c(\lambda;\underline{X}) = 0$.

While (2.2) is a desired candidate for testing positivity of state polynomials, one needs to consider bounded operators on an infinite-dimensional Hilbert space in order to obtain sums-of-squares certificates valid for sufficiently general C (cf. [HM04] for examples in the freely noncommutative setting, or the refutation of Connes' embedding conjecture in the tracial setting [JNV$^+$21, KS08]). Thus we mostly consider positivity on sets (2.1).

Remark 2.2. One might wonder why we restrict ourselves to real Hilbert spaces. In the complex framework, the only difference is that the symbol $\varsigma(w)$ needs to be split into two symbols, the real and imaginary part, to properly define $\varsigma(w^\star) = \overline{\varsigma(w)}$. Thus one is pressed to work with real variables also in the complex framework. Furthermore, every complex Hilbert space isometrically embeds into a real Hilbert space, meaning that constrained positivity over operators and states on a separable real Hilbert space also models positivity on a separable complex Hilbert space. The real framework is also more convenient to work with in optimization, especially from the perspective of implementation using the standard semidefinite programming solvers.

Remark 2.3. The correspondence between states and trace-class operators (or density matrices, in the finite-dimensional case) mentioned earlier begs the question why one cannot treat positivity and optimization of state polynomials with the corresponding already-established theory for trace polynomials [KMV22]. The reason is that optimization of trace polynomials pertains to evaluations of the *normalized* trace on matrices, or tracial states on von Neumann algebras. On the other hand, modelling state polynomials with trace polynomials (in an additional variable corresponding to the trace-class operator) would require considering the (usual) non-normalized trace. However, for evaluations of trace polynomials with respect to a non-normalized trace, there is no comparable dimension-independent theory of positivity and optimization.

3. A functional perspective on nc state polynomials

One can draw a comparison between nc state polynomials and noncommutative functions as developed e.g. in [KVV14]. While nc state polynomials are not noncommutative functions (since their evaluations on matrix tuples are not compatible with direct sums of matrix tuples), they nevertheless admit a closely related intrinsic characterization (Proposition 3.2), and like nc polynomials, they are determined on matrices of bounded size (Proposition 3.1). To establish these results, we recall the notion of trace polynomials [Pro76, KŠ17], originating from invariant theory. They are defined analogously as nc state polynomials, by adjoining commutative symbols $\operatorname{tr}(w)$ for $w \in \langle \underline{x} \rangle \setminus \{1\}$ to the free algebra $\mathbb{R}\langle \underline{x} \rangle$ in symmetric variables; the distinction is that in addition to $\operatorname{tr}(w) = \operatorname{tr}(w^*)$, these symbols satisfy the additional relation $\operatorname{tr}(uv) = \operatorname{tr}(vu)$.

The next proposition shows that nc state polynomials vanishing on all matrices of a given size need to have sufficiently high degree.

Proposition 3.1. *If $f \in \mathscr{S}$ is of degree d, then there exist $\lambda \in \mathcal{S}(\mathbb{R}^{2d+1})$ and $\underline{X} \in \mathrm{S}_{2d+1}(\mathbb{R})^n$ such that $f(\lambda; \underline{X}) \neq 0$.*

Proof. Let

$$f = \sum_{i=1}^{k} \alpha_i \prod_{j=1}^{m_i} \varsigma(u_{i,j}) v_i.$$

Define a trace polynomial in variables $x_0, \dots, x_n$

$$g = \sum_{i=1}^{k} \alpha_i (\operatorname{tr} x_0)^{d-m_i} \prod_{j=1}^{m_i} \operatorname{tr}(x_0 u_{i,j}) v_i.$$

Note that $g \neq 0$ since $f \neq 0$, and the degree of g (as a trace polynomial) is $2d$. By [Pro76, Proposition 8.3], g is not constantly zero on $\mathrm{S}_{2d+1}(\mathbb{R})^{n+1}$. Since positive definite matrices are Zariski dense in symmetric matrices, there exist $(P, \underline{X}) \in \mathrm{S}_{2d+1}(\mathbb{R}) \times \mathrm{S}_{2d+1}(\mathbb{R})^n$ with $P \succ 0$ such that $g(P, \underline{X}) \neq 0$. By the construction of g we have $g(\alpha P, \underline{X}) = \alpha^d g(P, \underline{X})$ for all $\alpha \in \mathbb{R}$. Therefore

$$f(\lambda; \underline{X}) = g\left(\tfrac{1}{\operatorname{tr} P} P, \underline{X}\right) \neq 0$$

where $\lambda \in \mathcal{S}(\mathbb{R}^{2d+1})$ is given by $\lambda(Y) = \operatorname{tr}(\frac{1}{\operatorname{tr} P} PY)$. □

As mentioned in Section 2.2, states in $\mathcal{S}(\mathbb{R}^k)$ are in one-to-one correspondence with $k \times k$ density matrices (positive semidefinite matrices in $\mathrm{S}_k(\mathbb{R})$ with trace 1); namely, a density matrix ρ gives rise to a state $Y \mapsto \operatorname{tr}(\rho Y)$. For the purpose of the next proposition we resort to this identification. Hence we view $\mathcal{S}(\mathbb{R}^k)$ as a Zariski dense subset of the affine space of symmetric matrices with trace 1. By a polynomial function on $\mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n$ we therefore refer to a polynomial function on the corresponding real affine space of dimension $\frac{k(k+1)}{2} - 1 + n\frac{k(k+1)}{2}$. Furthermore, $\mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n$ inherits the diagonal conjugate action of the orthogonal group $\mathbf{O}_k$ on tuples of symmetric $k \times k$ matrices:

$$O(X_0, \dots, X_n)O^* = (OX_0O^*, \dots, OX_nO^*).$$

The following is an nc state analog of [KVV14, Theorem 6.1] and [KŠ17, Proposition 3.1] for nc polynomials.

Proposition 3.2. *A sequence $(f_k)_{k \in \mathbb{N}}$ of polynomial maps $f_k : \mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n \to \mathrm{S}_k(\mathbb{R})$ satisfies*

(a) *$f_k(O(\lambda; \underline{X})O^*) = Of_k(\lambda; \underline{X})O^*$ for all $k \in \mathbb{N}$, $O \in \mathbf{O}_k$ and $(\lambda; \underline{X}) \in \mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n$,*

(b) *$f_{\ell k}(\rho \otimes \lambda; \underline{X}^{\oplus \ell}) = f_k(\lambda; \underline{X})^{\oplus \ell}$ for all $k, \ell \in \mathbb{N}$, $\rho \in \mathcal{S}(\mathbb{R}^\ell)$ and $(\lambda; \underline{X}) \in \mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n$,*

(c) *$\sup_k \deg f_k < \infty$,*

if and only if it is given by an nc state polynomial.

Proof. The implication ($\Leftarrow$) is routine, so we consider ($\Rightarrow$). By (a) and [Pro76, Theorem 7.3], for every $k \in \mathbb{N}$ there exists a trace polynomial T_k in variables $x_0, \dots, x_n$ such that T_k agrees with f_k on $\mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n$, the expression $\mathrm{tr}(x_0)$ does not appear in T_k (cf. [Pro76, Section 5]), and $\deg T_k = \deg f_k$. Denote $d = \max_k \deg f_k$.

For $k \geq d+1$, the trace polynomial T_k defined as above is unique [Pro76, Proposition 8.3]. Let

$$T_k = \sum_i \alpha_{k,i} \prod_j \mathrm{tr}(u_{i,j}) v_i \tag{3.1}$$

where $\prod_j \mathrm{tr}(u_{i,j}) v_i$ are distinct trace words with $|v_i| + \sum_j |u_{i,j}| \leq d$. By uniqueness and comparison of (b) for $\rho = \frac{1}{2} I_2$ and $\rho = \frac{1}{2}\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, the variable x_0 cannot appear in any v_i. Furthermore, if there are m occurrences of x_0 in $u_{i,j}$, then

$$\mathrm{tr}\Big(u_{i,j}\big(\rho \otimes \lambda; I \otimes X_1, \dots, I \otimes X_n\big)\Big) = \mathrm{tr}(\rho^m) \cdot \mathrm{tr}\big(u_{i,j}(\lambda; X_1, \dots, X_n)\big).$$

Therefore, uniqueness of T_k and comparison of (b) for $\rho = \frac{1}{2} I_2$ and $\rho = \frac{1}{4}\left(\begin{smallmatrix} 1 & 0 \\ 0 & 3 \end{smallmatrix}\right)$ imply that the variable x_0 can appear in each $u_{i,j}$ at most once. Finally, uniqueness and comparison of (b) for $\rho = \frac{1}{2} I_2$ and $\rho = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ show that x_0 appears in each $u_{i,j}$. Thus (3.1) becomes

$$T_k = \sum_i \alpha_{k,i} \prod_j \mathrm{tr}(x_0 u'_{i,j}) v'_i \tag{3.2}$$

where $u'_{i,j}, v'_i$ are words in $x_1, \dots, x_n$.

By the special structure (3.2),

$$T_{\ell k}(\lambda; \underline{X})^{\oplus \ell} = T_{\ell k}\left(\tfrac{1}{\ell} I_\ell \otimes \lambda, I_\ell \otimes \underline{X}\right) = f_{\ell k}\left(\tfrac{1}{\ell} I_\ell \otimes \lambda, I_\ell \otimes \underline{X}\right) = f_k(\lambda, \underline{X})^{\oplus \ell} = T_k(\lambda, \underline{X})^{\oplus \ell}$$

for all $\ell \in \mathbb{N}$, $k \geq d+1$ and $(\lambda, \underline{X}) \in \mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n$. By uniqueness of the T_k we thus have $T_{\ell k} = T_k$ for all $\ell \in \mathbb{N}$, $k \geq d+1$, and consequently $T_k = T_{d+1}$ for all $k \geq d+1$. The property (b) ensures that the evaluations of T_{d+1} and T_k for $k \leq d$ agree on $\mathrm{S}_k(\mathbb{R})^{n+1}$. Thus

$$f = \sum_i \alpha_{d+1,i} \prod_j \varsigma(u'_{i,j}) v'_i \in \mathscr{S}$$

is the desired nc state polynomial. $\square$

Corollary 3.3. *A sequence $(f_k)_{k \in \mathbb{N}}$ of polynomial maps $f_k : \mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n \to \mathbb{R}$ satisfies*

(a) *$f_k(O(\lambda; \underline{X})O^*) = f_k(\lambda; \underline{X})$ for all $k \in \mathbb{N}$, $O \in \mathbf{O}_k$ and $(\lambda; \underline{X}) \in \mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n$,*
(b) *$f_{\ell k}(\rho \otimes \lambda; \underline{X}^{\oplus \ell}) = f_k(\lambda; \underline{X})$ for all $k, \ell \in \mathbb{N}$, $\rho \in \mathcal{S}(\mathbb{R}^\ell)$ and $(\lambda; \underline{X}) \in \mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^n$,*
(c) *$\sup_k \deg f_k < \infty$,*

if and only if it is given by a state polynomial.

Proof. Replacing f_k with $\hat{f}_k := f_k I_k$, Proposition 3.2 implies $\hat{f}_k$ arises from an nc state polynomial, say $\hat{f}$. Add a new variable x_{n+1} and form the commutator $g := [\hat{f}, x_{n+1}]$. By assumption the image of $\hat{f}$ consists only of scalar matrices, so g always evaluates to 0 on all tuples $(\lambda;\underline{X}) \in \mathcal{S}(\mathbb{R}^k) \times \mathrm{S}_k(\mathbb{R})^{n+1}$. Thus by Proposition 3.1, $g = 0$. Hence $\hat{f}$ cannot have any "free" nc variables not bound by ς, i.e., $\hat{f} \in \mathscr{S}$. □

4. Hilbert's 17th problem for state polynomials

In this section we present a state analog of Artin's solution to the celebrated Hilbert's 17th problem. Let $\Omega \subseteq \mathscr{S}$ be the preordering [Mar08, Section 2.1] generated by $\{\varsigma(hh^\star) \colon h \in \boldsymbol{\mathscr{S}}\}$. That is, Ω is the smallest subset of $\mathscr{S}$ closed under sums and products that contains all squares and $\{\varsigma(hh^\star) \colon h \in \boldsymbol{\mathscr{S}}\}$. Hence Ω is the set of all sums of products of elements of the form $\varsigma(hh^\star)$ for $h \in \boldsymbol{\mathscr{S}}$ (since $a^2 = \varsigma(aa^\star)$ for $a \in \mathscr{S}$). Clearly, state polynomials in Ω are nonnegative on $\mathcal{D}_\emptyset$ and $\mathcal{D}_\emptyset^\infty$. Theorem 4.3 below shows that every state polynomial, nonnegative on $\mathcal{D}_\emptyset$, is a quotient of elements in Ω. This is in stark contrast with trace polynomials, where the analog of Theorem 4.3 only holds for one matrix variable [KPV21, KSV22].

For $d \in \mathbb{N}$ let $\langle \underline{x} \rangle_d$ denote the set of words with length at most d, and $D = |\langle \underline{x} \rangle_d| = \frac{n^{d+1}-1}{n-1}$. Let $H_d = (\varsigma(uv^\star))_{u,v\in\langle \underline{x} \rangle_d} \in \mathscr{S}^{D\times D}$. The following is a well-known statement adapted to the notation of this paper.

Lemma 4.1. *For each d there exists $(\lambda;\underline{X}) \in \vec{\mathcal{D}}_\emptyset$ such that $H_d(\lambda;\underline{X})$ is positive definite.*

Proof. By [HKM12, Lemma 3.2] there exists a unital $\star$-functional $L : \mathbb{R}\langle \underline{x} \rangle_{2d+2} \to \mathbb{R}$ such that $L(hh^\star) > 0$ for all nonzero $h \in \mathbb{R}\langle \underline{x} \rangle_{d+1}$. By [HKM12, Proposition 2.5], there is $(\lambda;\underline{X}) \in \vec{\mathcal{D}}_\emptyset$ such that $L(f) = \lambda(f(\underline{X}))$ for all $f \in \mathbb{R}\langle \underline{x} \rangle_{2d}$. Then $H_d(\lambda;\underline{X})$ is positive definite by construction. □

Proposition 4.2. *Every principal minor of H_d is a quotient of two elements in Ω.*

Proof. Let Ξ be the generic $D\times D$ symmetric matrix; that is, the entries of Ξ are commuting indeterminates, related only by Ξ being symmetric. Let A be the real polynomial algebra generated by the entries of Ξ, and $T \subseteq A$ its real subalgebra generated by $\{\mathrm{tr}(\Xi^j) \colon 1 \le j \le D\}$. Let P be the preordering in T generated by

$$\{\mathrm{tr}(h(\Xi)^2),\ \mathrm{tr}(h(\Xi)\cdot \Xi \cdot h(\Xi)) \colon h \in T[\Xi]\}.$$

Let $m \in T$ be an arbitrary principal minor of Ξ. By [KŠV18, Lemma 4.1 and Theorem 4.13] there exist $p, q \in P$ and $k \in \mathbb{N}$ such that

$$qm = m^{2k} + p. \tag{4.1}$$

Let $\mathbf{w}$ be the vector of words in $\langle \underline{x} \rangle_d$ (ordered degree-lexicographically); then H_d is obtained by applying ς to $\mathbf{w}\mathbf{w}^\star \in \mathbb{R}\langle \underline{x} \rangle^{D\times D}$ entry-wise. If $h \in T[\Xi] \subseteq A^{d\times d}$, then $h(H_d) \in \mathscr{S}^{D\times D}$ and $h(H_d)\mathbf{w} \in \boldsymbol{\mathscr{S}}^D$; moreover,

$$\begin{aligned} &\mathrm{tr}(h(\Xi)^2) \in \Omega, \\ &\mathrm{tr}\left(h(H_d)\, H_d\, h(H_d)\right) = \sum_{j=1}^{D} \varsigma\big((h(H_d)\mathbf{w})_j\, (h(H_d)\mathbf{w})_j^\star \big) \in \Omega. \end{aligned} \tag{4.2}$$

If m', p', q' are obtained from m, p, q by replacing Ξ with H_d, then $p', q' \in \Omega$ by (4.2). Furthermore, $q' \neq 0$ since the right-hand side of (4.1) is strictly positive when evaluated at a positive definite state evaluation of H_d, which exists by Lemma 4.1. Therefore m', a principal minor of H_d, is a quotient of elements in Ω. □

The following is a solution to Hilbert's 17th problem for state polynomials.

Theorem 4.3. *The following are equivalent for $a \in \mathscr{S}$:*

(i) *$a(\lambda; \underline{X}) \geq 0$ for all $\underline{X} \in \mathrm{S}_K(\mathbb{R})^n$ and vector states $\lambda \in \mathcal{S}(\mathbb{R}^K)$, where $K = \frac{1}{n-1}(n^{\lceil \frac{1+\deg a}{2} \rceil} - 1)$;*

(ii) *a is nonnegative on $\mathcal{D}_\emptyset^\infty$;*

(iii) *a is a quotient of two elements in Ω.*

Proof. (ii)⇒(i) Clear.

(i)⇒(iii) Suppose a is not a quotient of elements in Ω. Let $d = \lceil \frac{1+\deg a}{2} \rceil$ and $R = \mathbb{R}[\varsigma(w) \colon w \in \langle \underline{x} \rangle_{2d} \setminus \{1\}]$. Then $a \in R$, and R is a finitely generated polynomial ring. Let $M \subseteq R$ be the set of all principal minors of H_d. By Proposition 4.2, a is not a quotient of elements in the preordering in R generated by M. By the Krivine–Stengle Positivstellensatz [Mar08, Theorem 2.2.1] there exists a homomorphism $\varphi : R \to \mathbb{R}$ such that $\varphi(a) < 0$ and $\varphi(M) \subseteq \mathbb{R}_{\geq 0}$. Applying φ entry-wise to H_d therefore results in a positive semidefinite matrix. Define $L : \mathbb{R}\langle \underline{x} \rangle_{2d} \to \mathbb{R}$ as $L(f) = \varphi(\varsigma(f))$. Then L is a unital $\star$-functional, and $L(gg^\star) \geq 0$ for $g \in \mathbb{R}\langle \underline{x} \rangle_d$. By Lemma 4.1 there exists a unital $\star$-functional $L' : \mathbb{R}\langle \underline{x} \rangle_{2d} \to \mathbb{R}$ such that $L'(gg^\star) > 0$ for nonzero $g \in \mathbb{R}\langle \underline{x} \rangle_d$. For every $\varepsilon \in (0,1)$ denote the unital $\star$-functional $L_\varepsilon = (1-\varepsilon)L + \varepsilon L'$; then $L_\varepsilon(gg^\star) > 0$ for all nonzero $g \in \mathbb{R}\langle \underline{x} \rangle_d$. By [HKM12, Proposition 2.5] there exists $(\lambda_\varepsilon; \underline{X}_\varepsilon) \in \mathcal{S}(\mathbb{R}^K) \times \mathrm{S}_K(\mathbb{R})^n$, with $K = \dim \mathbb{R}\langle \underline{x} \rangle_{d-1} = \frac{n^d - 1}{n-1}$ and λ_ε a vector state, such that $L_\varepsilon(f) = \lambda_\varepsilon(f(\underline{X}_\varepsilon))$ for all $f \in \mathbb{R}\langle \underline{x} \rangle_{2d-1}$. Then

$$\lim_{\varepsilon \to 0} a(\lambda_\varepsilon; \underline{X}_\varepsilon) = a(\lambda_0; \underline{X}_0) = \varphi(a) < 0$$

and so $a(\lambda_\varepsilon; \underline{X}_\varepsilon) < 0$ for some $\varepsilon \in (0,1)$.

(iii)⇒(ii) Let $a = p/q$ for some $p, q \in \Omega$ with $q \neq 0$. Clearly p and q are nonnegative on $\mathcal{D}_\emptyset^\infty$. Let $\underline{X}$ be a tuple of self-adjoint bounded operators on an infinite-dimensional separable Hilbert space $\mathcal{H}$, and let λ be a state on $\mathcal{B}(\mathcal{H})$. Suppose $a(\lambda; \underline{X}) < 0$. Since $q \neq 0$, by Proposition 3.1 there exist $\underline{Y} \in \mathrm{S}_k(\mathbb{R})^n$ and a $\mu \in \mathcal{S}(\mathbb{R}^k)$ such that $q(\mu; \underline{Y}) \neq 0$. Let $\iota : \mathbb{R}^k \to \mathcal{H}$ be an isometry, and define $X'_j = \iota \circ Y_j \circ \iota^*$ and $\lambda' = \mu \circ \iota^*$. Then $q(\lambda'; \underline{X}') \neq 0$. For $\varepsilon \in (0,1)$ set $\lambda_\varepsilon = (1-\varepsilon)\lambda + \varepsilon\lambda'$ and $\underline{X}_\varepsilon = (1-\varepsilon)\underline{X} + \varepsilon\underline{X}'$. Since $q(\lambda_1; \underline{X}_1) \neq 0$ and q is a polynomial in state symbols, we have $q(\lambda_\varepsilon; \underline{X}_\varepsilon) \neq 0$ for all but finitely many $\varepsilon \in (0,1)$. On the other hand, $\lim_{\varepsilon \to 0} a(\lambda_\varepsilon; \underline{X}_\varepsilon) = a(\lambda_0; \underline{X}_0) < 0$. Therefore there exists $\varepsilon \in (0,1)$ such that $q(\lambda_\varepsilon; \underline{X}_\varepsilon) \neq 0$ and $a(\lambda_\varepsilon; \underline{X}_\varepsilon) < 0$, so

$$0 > a(\lambda_\varepsilon; \underline{X}_\varepsilon) = \frac{p(\lambda_\varepsilon; \underline{X}_\varepsilon)}{q(\lambda_\varepsilon; \underline{X}_\varepsilon)} \geq 0,$$

a contradiction. □

Theorem 4.3 shows that *every* state polynomial inequality valid on all matrices is an algebraic consequence of states on hermitian squares, and is moreover also valid on all operators. A well-known example is the Cauchy-Schwarz inequality, which admits an algebraic certificate

$$\varsigma(x_1^2)\varsigma(x_2^2)-\varsigma(x_1x_2)^2=\frac{\varsigma\Big(\big(\varsigma(x_1^2)x_2-\varsigma(x_1x_2)x_1\big)^2\Big)}{\varsigma(x_1)^2}.$$

Alternatively, one can recognize $\varsigma(x_1^2)\varsigma(x_2^2)-\varsigma(x_1x_2)^2$ as the 2×2 principal minor of H_1 indexed by x_1, x_2.

Example 4.4. A less evident example of a globally nonnegative state polynomial is

$$\begin{aligned} a=&\big(\varsigma(x_1^2)\varsigma(x_2^2)-\varsigma(x_1x_2)^2\big)\varsigma(x_2x_1^2x_2)+2\varsigma(x_1x_2)\varsigma(x_2x_1x_2)\varsigma(x_1^2x_2)\\ &-\varsigma(x_1^2)\varsigma(x_2x_1x_2)^2-\varsigma(x_2^2)\varsigma(x_1^2x_2)^2. \end{aligned}$$

There are two ways to certify nonnegativity of a. Consider a principal submatrix of H_2,

$$s=\begin{pmatrix}\varsigma(x_1^2) & \varsigma(x_1x_2) & \varsigma(x_1^2x_2)\\ \varsigma(x_1x_2) & \varsigma(x_2^2) & \varsigma(x_2x_1x_2)\\ \varsigma(x_1^2x_2) & \varsigma(x_2x_1x_2) & \varsigma(x_2x_1^2x_2)\end{pmatrix}\in\mathscr{S}^{3\times 3}.$$

Then $a=\det(s)$, which gives the first cue for a being nonnegative. To obtain a certificate in terms of Theorem 4.3, let

$$\sigma_2=\varsigma(x_1^2)\varsigma(x_2x_1^2x_2)-\varsigma(x_1^2x_2)^2+\varsigma(x_1^2)\varsigma(x_2^2)-\varsigma(x_1x_2)^2+\varsigma(x_2^2)\varsigma(x_2x_1^2x_2)-\varsigma(x_2x_1x_2)^2$$

which is one of the coefficients of the characteristic polynomial of s. Note that σ_2 is the sum of three quotients of elements in Ω, by the argument used above for the Cauchy-Schwarz inequality. If

$$\begin{pmatrix}h_1\\ h_2\\ h_3\end{pmatrix}=(s^2-\operatorname{tr}(s)s+\sigma_2I_3)\cdot\begin{pmatrix}x_1\\ x_2\\ x_2x_1\end{pmatrix}\in\boldsymbol{\mathscr{S}}^3$$

then a direct calculation shows that

$$a=\frac{\varsigma(h_1h_1^\star)+\varsigma(h_2h_2^\star)+\varsigma(h_3h_3^\star)}{\sigma_2},$$

so a is a quotient of elements in Ω. The choice of h_j is inspired by the proof of Proposition 4.2 and [KŠV18, Example 6.1].

5. Archimedean Positivstellensatz for state polynomials

In this section we give a version of Putinar's Positivstellensatz [Put93] for state polynomials subject to archimedean constraints, Theorem 5.5, which is later applied to state polynomial optimization in Section 6. First we address which functionals $\mathbb{R}\langle \underline{x}\rangle\to\mathbb{R}$ are given by states and evaluations on tuples of bounded operators. The following is a variant of the well-known Gelfand-Naimark-Segal (GNS) construction.

Proposition 5.1. *Let $L : \mathbb{R}\langle \underline{x} \rangle \to \mathbb{R}$ be a unital $\star$-functional. If*

(a) $L(pp^\star) \geq 0$ *for all* $p \in \mathbb{R}\langle \underline{x} \rangle$, *and*
(b) *there is* $N > 0$ *such that* $L(p(N - x_1^2 - \cdots - x_n^2)p^\star) \geq 0$ *for all* $p \in \mathbb{R}\langle \underline{x} \rangle$,

then there exist a vector state $\lambda \in \mathcal{S}(\mathcal{H})$ *and a tuple of self-adjoint operators* $\underline{X} \in \mathcal{B}(\mathcal{H})^n$ *such that* $L(p) = \lambda(p(\underline{X}))$ *for all* $p \in \mathbb{R}\langle \underline{x} \rangle$.

Proof. Apply [BKP16, Theorem 1.27] to the quadratic module in $\mathbb{R}\langle \underline{x} \rangle$ generated by $\{N - x_1^2 - \cdots - x_n^2\}$. $\square$

Remark 5.2. It is easy to see that (b) in Proposition 5.1 can be replaced by

(b') there is $N > 0$ such that $L(ww^\star) \leq N^{|w|}$ for all $w \in \langle \underline{x} \rangle$.

On the other hand, it cannot be replaced by

(b") there is $N > 0$ such that $L(x_j^{2k}) \leq N^k$ for all $j = 1, \ldots, n$ and $k \in \mathbb{N}$,

as in the tracial setup [KMV22, Proposition 3.2]. Namely, for non-tracial states not all mixed moments can be bounded with univariate moments. For example, let

$$v = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad X_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & \gamma \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma \\ 0 & 0 & 1 & 0 \\ 0 & \gamma & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Then $\langle X_j^{2k} v, v \rangle = 1$ for all j, k but $\langle X_1 X_2 X_3 v, v \rangle = \gamma$ can be arbitrarily large.

Due to Proposition 5.1, we focus on state polynomial positivity subject to balanced constraint sets with the following property. We say that $C \subseteq \mathscr{S}$ is *algebraically bounded* if there is $N > 0$ such that

$$N - x_1^2 - \cdots - x_n^2 = \sum_i p_i c_i p_i^\star$$

for some $c_i \in \{1\} \cup C \cap \mathbb{R}\langle \underline{x} \rangle$ and $p_i \in \mathbb{R}\langle \underline{x} \rangle$ (in other words, $C \cap \mathbb{R}\langle \underline{x} \rangle$ generates an archimedean quadratic module in $\mathbb{R}\langle \underline{x} \rangle$).

Next we turn to a notion from real algebra [Mar08, Sch09]. A subset $\mathcal{M} \subseteq \mathscr{S}$ is called a quadratic module if $1 \in \mathcal{M}$, $\mathcal{M} + \mathcal{M} \subseteq \mathcal{M}$ and $a^2\mathcal{M}$ for all $a \in \mathscr{S}$. For $M \subseteq \mathscr{S}$ let $\mathrm{QM}(M)$ denote the quadratic module generated by M. Given a quadratic module $\mathcal{M} \subseteq \mathscr{S}$ that is archimedean (i.e., for each $f \in \mathscr{S}$ there is $m > 0$ such that $m \pm f \in \mathcal{M}$), we consider the real points of the real spectrum $\mathrm{Sper}_\mathcal{M} \mathscr{S}$, namely the set $\chi_\mathcal{M}$ defined by

$$\chi_\mathcal{M} := \{\varphi : \mathscr{S} \to \mathbb{R} \colon \varphi \text{ homomorphism}, \ \varphi(\mathcal{M}) \subseteq \mathbb{R}_{\geq 0}, \ \varphi(1) = 1\}. \tag{5.1}$$

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

Proposition 5.3. *Let $\mathcal{M} \subseteq \mathscr{S}$ be an archimedean quadratic module. Then, for all $a \in \mathscr{S}$, one has*

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

Since the algebra $\mathscr{S}$ is not finitely generated, it is in general not straightforward to determine if a quadratic module in $\mathscr{S}$ is archimedean. Nevertheless, the next lemma shows that quadratic modules arising from algebraically bounded sets are archimedean. To $C \subseteq \boldsymbol{\mathscr{S}}$ we assign

$$C^\varsigma := \{\varsigma(pcp^\star) \colon p \in \mathbb{R}\langle \underline{x} \rangle, c \in \{1\} \cup C\} \subseteq \mathscr{S}.$$

Lemma 5.4. *If C is balanced and algebraically bounded then the quadratic module $\mathrm{QM}(C^\varsigma) \subseteq \mathscr{S}$ is archimedean.*

Proof. It suffices to show that the generators of $\mathscr{S}$ are bounded with respect to $\mathrm{QM}(C^\varsigma)$ [Cim09], i.e., that for every $w \in \langle \underline{x} \rangle$ there exists $m > 0$ such that $m \pm \varsigma(w) \in \mathrm{QM}(C^\varsigma)$. Since

$$N\varsigma(ww^\star) - \varsigma(wx_j^2 w^\star) = \varsigma\left(w(N - x_j^2)w^\star\right) \in \mathrm{QM}(C^\varsigma),$$

the induction on $|w|$ implies that for every w there is $m' > 0$ such that $m' - \varsigma(ww^\star) \in \mathrm{QM}(C^\varsigma)$. Then

$$\tfrac{1}{4} + m' \pm \varsigma(w) = \varsigma\left((\tfrac{1}{2} \pm w)(\tfrac{1}{2} \pm w)^\star\right) + m' - \varsigma(ww^\star) \in \mathrm{QM}(C^\varsigma). \qquad \square$$

We are now ready to prove an analog of the noncommutative Helton-McCullough Positivstellensatz [HM04] for state polynomials subject to nc state constraints.

Theorem 5.5. *Let $C \subseteq \boldsymbol{\mathscr{S}}$ be balanced and algebraically bounded. Then for $a \in \mathscr{S}$ the following are equivalent:*

(i) *$a(\lambda; \underline{X}) \geq 0$ for all $(\lambda; \underline{X}) \in \vec{\mathcal{D}}_C^\infty$;*
(ii) *$a(\lambda; \underline{X}) \geq 0$ for all $(\lambda; \underline{X}) \in \mathcal{D}_C^\infty$;*
(iii) *$a + \varepsilon \in \mathrm{QM}(C^\varsigma)$ for all $\varepsilon > 0$.*

Proof. (iii)⇒(ii) If $(\lambda; \underline{X}) \in \mathcal{D}_C^\infty$, then

$$c(\lambda; \underline{X}) \succeq 0, \qquad \lambda p(\underline{X}) p(\underline{X})^* \succeq 0$$

for all $c \in C$ and $p \in \mathbb{R}\langle \underline{x} \rangle$, hence $s(\lambda; \underline{X}) \geq 0$ for all $s \in \mathrm{QM}(C^\varsigma)$, and so $a(\lambda; \underline{X}) \geq 0$.

(ii)⇒(i) Clear.

(i)⇒(iii) Suppose $a + \varepsilon \notin \mathrm{QM}(C^\varsigma)$ for some $\varepsilon > 0$. By Proposition 5.3 there exists a unital homomorphism $\varphi : \mathscr{S} \to \mathbb{R}$ with $\varphi(\mathrm{QM}(C^\varsigma)) \subseteq \mathbb{R}_{\geq 0}$ and $\varphi(a) < 0$. Hence

$$\varphi(\varsigma(pp^\star)) \geq 0, \qquad \varphi(\varsigma(p(N - x_1^2 - \cdots - x_n^2)p^\star)) \geq 0$$

for all $p \in \mathbb{R}\langle \underline{x} \rangle$. Consider the unital $\star$-functional $L : \mathbb{R}\langle \underline{x} \rangle \to \mathbb{R}$ given by $L(p) = \varphi(\varsigma(p))$. By Proposition 5.1, there exist a vector state $\lambda \in \mathcal{S}(\mathcal{H})$ and $\underline{X} = \underline{X}^* \in \mathcal{B}(\mathcal{H})^n$ such that $L(p) = \lambda(p(\underline{X}))$ for all $p \in \mathbb{R}\langle \underline{x} \rangle$. Therefore $\varphi(b) = b(\lambda; \underline{X})$ for all $b \in \mathscr{S}$. Let $v \in \mathcal{H}$ be a unit vector such that $\lambda(Y) = \langle Yv, v \rangle$ for all $Y \in \mathcal{B}(\mathcal{H})$, and $P \in \mathcal{B}(\mathcal{H})$ be the orthogonal projection onto $\overline{\{p(\underline{X})v \colon p \in \mathbb{R}\langle \underline{x} \rangle\}}$. Then $\langle q(\underline{X})v, v \rangle = \langle q(P\underline{X}P)v, v \rangle$ for all $q \in \mathbb{R}\langle \underline{x} \rangle$.

Thus we can without loss of generality assume $PX_j = X_j$; that is, the operators X_j can be replaced by their compressions PX_jP. Hence

$$\langle c(\lambda;\underline{X})p(\underline{X})v, p(\underline{X})v\rangle = \varphi(\varsigma(p^\star cp)) \geq 0 \qquad \text{for } c\in C,\ p\in\mathbb{R}\langle\underline{x}\rangle$$

together with $(I-P)X_j = 0$ implies $c(\lambda;\underline{X})\succeq 0$, and so $(\lambda,\underline{X})\in\vec{\mathcal{D}}_C^\infty$. Finally, $a(\lambda;\underline{X}) = \varphi(a) < 0$. □

Given $a\in\mathscr{S}$ and $c,p\in\boldsymbol{\mathscr{S}}$ we have $a^2\varsigma(pcp^\star) = \varsigma((ap)c(ap)^\star)$. Therefore

$$\mathrm{QM}(C^\varsigma) \subseteq \left\{\sum_{i=1}^K \varsigma(f_ic_if_i^\star)\colon K\in\mathbb{N},\ f_i\in\boldsymbol{\mathscr{S}},\ c_i\in\{1\}\cup C\right\}, \tag{5.2}$$

an inclusion we shall make use of in Sections 5.1 and 6.1 below.

5.1. **Quadratic state modules.** In this section we provide an alternative form of the Positivstellensatz for state polynomials that underlines their noncommutative roots. A subset $\mathcal{M}\subseteq\boldsymbol{\mathscr{S}}$ is called a *quadratic state module* if

$$1\in\mathcal{M},\ \mathcal{M}+\mathcal{M}\subseteq\mathcal{M},\ f\mathcal{M}f^\star\subseteq\mathcal{M}\ \forall f\in\boldsymbol{\mathscr{S}},\ \varsigma(\mathcal{M})\subseteq\mathcal{M}.$$

Given $C\subseteq\boldsymbol{\mathscr{S}}$ let $\mathrm{QM}_\varsigma(C)$ be the quadratic state module generated by C, i.e., the smallest quadratic state module in $\boldsymbol{\mathscr{S}}$ containing C. We start with an alternative description of quadratic state modules.

Lemma 5.6. *Let $C\subseteq\boldsymbol{\mathscr{S}}$.*

(1) *Elements of $\mathrm{QM}_\varsigma(\emptyset)$ are precisely sums of*

$$\varsigma(h_1h_1^\star)\cdots\varsigma(h_\ell h_\ell^\star)h_0h_0^\star$$

for $h_i\in\boldsymbol{\mathscr{S}}$.

(2) *Elements of $\mathrm{QM}_\varsigma(C)$ are precisely sums of*

$$q_1,\quad h_1c_1h_1^\star,\quad \varsigma(h_2c_2h_2^\star)q_2$$

for $h_i\in\boldsymbol{\mathscr{S}}$, $q_i\in\mathrm{QM}_\varsigma(\emptyset)$, $c_i\in C$.

(3) *Elements of $\varsigma(\mathrm{QM}_\varsigma(C)) = \mathrm{QM}_\varsigma(C)\cap\mathscr{S}$ are precisely sums of*

$$\varsigma(h_1h_1^\star)\cdots\varsigma(h_\ell h_\ell^\star)\varsigma(h_0ch_0^\star)$$

for $h_i\in\boldsymbol{\mathscr{S}}$ and $c\in C\cup\{1\}$.

In particular, $\varsigma(\mathrm{QM}_\varsigma(\emptyset))$ is the preordering Ω from Section 4.

Proof. Straightforward. □

Note that while $\varsigma(\mathrm{QM}_\varsigma(\emptyset))$ is a preordering in $\mathscr{S}$, this is not necessarily the case for $\varsigma(\mathrm{QM}_\varsigma(C))$ in general (namely, $\varsigma(\mathrm{QM}_\varsigma(C))$ does not need to contain elements of the form $\varsigma(h_1c_1h_1^\star)\varsigma(h_2c_2h_2^\star)$ for $c_i\in C$). A quadratic state module $\mathcal{M}$ is called *archimedean* if for every $f\in\mathrm{Sym}\,\boldsymbol{\mathscr{S}}$ there exists $N>0$ such that $N\pm f\in\mathcal{M}$ (note that even though $\mathcal{M}$ might not be contained in $\mathrm{Sym}\,\boldsymbol{\mathscr{S}}$, we only consider symmetric f).

Proposition 5.7. *A quadratic state module $\mathcal{M}$ is archimedean if and only if there exists $N > 0$ such that $N - x_1^2 - \cdots - x_n^2 \in \mathcal{M}$.*

Proof. The forward implication is obvious. For the converse, note that the set $\mathcal{M} \cap \mathbb{R}\langle \underline{x} \rangle$ is an archimedean quadratic module. Thus, for all $p = p^\star \in \mathbb{R}\langle \underline{x} \rangle$ there exists $M > 0$ such that

$$M \pm p \in \mathcal{M} \cap \mathbb{R}\langle \underline{x} \rangle \,. \tag{5.3}$$

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

$$H = \{h \in \mathscr{S} \colon \exists M > 0 \text{ s.t. } M - hh^\star \in \mathcal{M}\}\,,$$

is closed under involution, addition, subtraction and multiplication, i.e., is a $\star$-subalgebra of $\mathscr{S}$ [Cim09]. A symmetric element $f \in \mathscr{S}$ is in H if and only if there is some $M > 0$ with $M \pm f \in \mathcal{M}$.

For every $w \in \langle \underline{x} \rangle$ we have

$$\varsigma(ww^\star) - \varsigma(w)^2 = \varsigma\big((w - \varsigma(w))(w - \varsigma(w))^\star\big) \in \mathcal{M}. \tag{5.4}$$

By (5.3) and the fact that $\mathcal{M}$ is preserved under ς, there is some $M > 0$ with $M - \varsigma(ww^\star) \in \mathcal{M}$. Adding this to (5.4) yields $M - \varsigma(w)^2 \in \mathcal{M}$. Thus, by the definition of H, $\varsigma(w) \in \mathcal{M}$. The desired result now follows since H is a subalgebra of $\mathscr{S}$. □

The following is the state version of Theorem 5.5. Note that while the constraints in Corollary 5.8 are nc state polynomials, the objective function needs to be a state polynomial (cf. [KMV22, Example 4.6]).

Corollary 5.8. *Let $\mathcal{M} \subseteq \mathscr{S}$ be an archimedean quadratic state module and $a \in \mathscr{S}$. The following are equivalent:*

(i) *$a(\lambda; \underline{X}) \geq 0$ for all $(\lambda, \underline{X}) \in \vec{\mathcal{D}}^\infty_{\mathcal{M}}$;*
(ii) *$a(\lambda; \underline{X}) \geq 0$ for all $(\lambda, \underline{X}) \in \mathcal{D}^\infty_{\mathcal{M}}$;*
(iii) *$a + \varepsilon \in \mathcal{M}$ for all $\varepsilon > 0$.*

Proof. By Proposition 5.7, there exists $N > 0$ such that $N - x_1^2 - \cdots - x_n^2 \in \mathcal{M}$. Furthermore, $\mathrm{QM}(\mathcal{M}^\varsigma) \subseteq \mathcal{M}$ by (5.2). Therefore (i)⇒(iii) follows from Theorem 5.5, and (iii)⇒(ii)⇒(i) is clear. □

5.2. **Nonexistence of a Krivine-Stengle or Schmüdgen Positivstellensatz for state polynomials.** We finish this section by commenting on two classical non-archimedean Positivstellensätze from real algebraic geometry, and how their straightforward (albeit possibly naïve) analogs for state polynomials fail.

A quadratic state module $\mathcal{P} \subseteq \mathscr{S}$ is a *state preordering* if $\varsigma(\mathcal{P}) \cdot \mathcal{P} \subseteq \mathcal{P}$. Note that $\varsigma(\mathcal{P}) = \mathcal{P} \cap \mathscr{S}$ is then a preordering (in the usual sense) in $\mathscr{S}$. Moreover, if $\mathcal{P}$ is a state preordering generated by C, then $\varsigma(\mathcal{P})$ is the preordering generated by $\varsigma(hch^\star)$ for $c \in \{1\} \cup C$.

Consider the following two statements about a finitely generated state preordering $\mathcal{P}$ and $a \in \mathscr{S}$, which would be analogs of the Krivine-Stengle Positivstellensatz [Mar08, Theorem 2.2.1] and the Schmüdgen Positivstellensatz [Mar08, Corollary 6.1.2], respectively:

(A) if $a|_{\mathcal{D}_{\mathcal{P}}^\infty} \geq 0$ then there exist $p_1, p_2 \in \mathcal{P}$ and $k \in \mathbb{N}$ such that $ap_1 = a^{2k} + p_2$;

(B) if $\mathcal{D}_{\mathcal{P}}^\infty$ is bounded in operator norm and $a|_{\mathcal{D}_{\mathcal{P}}^\infty} \geq \varepsilon$ for some $\varepsilon > 0$ then $a \in \mathcal{P}$.

Example 5.9. Let $\mathcal{P}$ be the state preordering generated by

$$\Big\{ \pm\left(1 + [x_1, x_2]^2\right), \pm\big[[x_1, x_2], x_1\big], \pm\big[[x_1, x_2], x_2\big]\Big\},$$

and $a = -\varsigma(x_1)$. Then

(1) $\mathcal{D}_{\mathcal{P}}^\infty = \emptyset$;

(2) there is a homomorphism $\varphi : \mathscr{S} \to \mathbb{R}$ such that $\varphi(\varsigma(\mathcal{P})) = \mathbb{R}_{\geq 0}$ and $\varphi(a) < 0$;

(3) The above implications (A) and (B) both fail.

Proof. (1) Let $\mathcal{H}$ be a complex Hilbert space, and suppose there exist $X_1, X_2 \in \mathcal{B}(\mathcal{H})$ such that $[X_1, X_2]$ commutes with X_1, X_2 and

$$I + [X_1, X_2]^2 = 0. \tag{5.5}$$

By the GNS construction (applied with any state on $\mathcal{B}(\mathcal{H})$) one can then without loss of generality assume that there exists $v \in \mathcal{H}$ such that $\{p(\underline{X})v : p \in \mathbb{R}\langle \underline{x} \rangle\}$ is dense in $\mathcal{H}$. Then $[X_1, X_2]$ is central in $\mathcal{B}(\mathcal{H})$, and so (5.5) implies $[X_1, X_2] = \pm iI$. But this contradicts nonexistence of bounded representations of the Weyl algebra. Therefore $\mathcal{D}_{\mathcal{P}}^\infty = \emptyset$.

(2) The standard (Bargmann-Fock) unbounded $\star$-representation of the Weyl algebra is given by the unbounded operator T acting on the complex Hilbert space ℓ^2 with the orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$ as $Te_n = \sqrt{n}e_{n+1}$. The domain of this representation contains $\bigoplus_{n\in\mathbb{N}} \mathbb{C}e_n$, and T satisfies $T^*T - TT^* = I$. Let $v = \frac{1}{\sqrt{2}}(e_1 + e_2)$ and $X_1 = \frac{1}{2}(T + T^*)$, $X_2 = \frac{1}{2i}(T - T^*)$. Then X_1, X_2 are self-adjoint unbounded operators, and $[X_1, X_2] = iI$. Define a homomorphism $\varphi : \boldsymbol{\mathscr{S}} \to \mathbb{R}$ by

$$\varphi(\varsigma(w)) = \operatorname{Re} \langle w(X_1, X_2)v, v\rangle$$

for $w \in \langle \underline{x} \rangle$. By the construction, $\varphi(\varsigma(\mathcal{P})) = \mathbb{R}_{\geq 0}$ and $\varphi(a) = -\frac{1}{2}$.

(3) Note that $\mathcal{D}_{\mathcal{P}}^\infty$ is bounded and $a|_{\mathcal{D}_{\mathcal{P}}^\infty} \geq 1$ vacuously by (1). Then (B) fails since $a \notin \varsigma(\mathcal{P}) = \mathcal{P} \cap \mathscr{S}$ by (2). If $ap_1 = a^{2k} + p_2$ for $p_1, p_2 \in \mathcal{P}$, then

$$0 \geq \varphi(a)\varphi(\varsigma(p_1)) = \varphi(a)^{2k} + \varphi(\varsigma(p_2)) > 0,$$

a contradiction; thus also (A) fails. □

6. Optimization of state polynomials

In this section we apply Theorem 5.5 to optimization of state objective functions subject to nc state constraints. Doing so, we obtain a converging hierarchy of SDP relaxations in Section 6.1, and its dual in Section 6.2. When flatness occurs in the latter hierarchy, one can extract a finite-dimensional minimizer as shown in Section 6.3. Finally, Section 6.4 presents sparsity-based approaches to reducing the sizes of the constructed SDP hierarchies.

Recall that the degree of a $\boldsymbol{\mathscr{S}}$-word $\prod_j \varsigma(u_j)v$ equals $|v| + \sum_j |u_j|$, and the degree of $f \in \boldsymbol{\mathscr{S}}$ is the maximal degree of $\boldsymbol{\mathscr{S}}$-words in the expansion of f. Let $\boldsymbol{\mathscr{S}}_d$ be subspace of nc state polynomials of degree at most d, and denote $\Delta(n,d) = \dim \boldsymbol{\mathscr{S}}_d$. It is not hard to estimate $n^d \le \Delta(n,d) \le (2n)^{d+1}$. Also, let $\mathscr{S}_d = \boldsymbol{\mathscr{S}}_d \cap \mathscr{S}$.

Throughout the rest of the paper we restrict ourselves to constraint sets $C \subseteq \boldsymbol{\mathscr{S}}$ such that $C \cap \boldsymbol{\mathscr{S}}_d$ is finite for all $d \in \mathbb{N}$. In polynomial optimization one typically focuses on finite sets of constraints, but this slightly more general setup is needed later in Section 8.

6.1. SDP hierarchy for state polynomial optimization.

For a balanced $C \subseteq \boldsymbol{\mathscr{S}}$ and $d \in \mathbb{N}$ define

$$\mathcal{M}(C)_d := \left\{ \sum_{i=1}^{K} \varsigma(f_i c_i f_i^\star) \colon K \in \mathbb{N},\ f_i \in \boldsymbol{\mathscr{S}},\ c_i \in \{1\} \cup C,\ \deg(f_i c_i f_i^\star) \le 2d \right\}. \tag{6.1}$$

By (6.1) it is clear that checking membership in $\mathcal{M}(C)_d$ can be formulated as an SDP. Furthermore, $\mathcal{M}(C)_d$ for $d = 1, 2, \ldots$ is an increasing sequence of convex cones whose union by (5.2) contains the quadratic module $\mathrm{QM}(C^\varsigma)$ from Section 5.

Given a state polynomial $a \in \mathscr{S}$, one can use $\mathcal{M}(C)_d$ to design a hierarchy of SDP relaxations for minimizing a over the state semialgebraic set $\mathcal{D}_C^\infty$. Let us define $a_{\min}$ and $a_{\min}^\infty$ as follows:

$$a_{\min} := \inf\{a(\lambda; \underline{X}) \colon (\lambda, \underline{X}) \in \mathcal{D}_C\}, \tag{6.2}$$

$$a_{\min}^\infty := \inf\{a(\lambda; \underline{X}) \colon (\lambda, \underline{X}) \in \mathcal{D}_C^\infty\}. \tag{6.3}$$

Since $\mathcal{D}_C$ embeds into $\mathcal{D}_C^\infty$, one has $a_{\min}^\infty \le a_{\min}$. One can approximate $a_{\min}^\infty$ from below via the following hierarchy of SDPs, indexed by $d \ge \frac{\deg a}{2}$:

$$a_{\min,d} = \sup\{m \colon a - m \in \mathcal{M}(C)_d\}. \tag{6.4}$$

Corollary 6.1. *If C is balanced and algebraically bounded, the SDP hierarchy (6.4) provides a sequence of lower bounds $(a_{\min,d})_{d \ge \frac{\deg a}{2}}$ monotonically converging to $a_{\min}^\infty$.*

Proof. By (6.1) and (6.4) and it is clear that $a_{\min,d} \le a_{\min}^\infty$. As $\mathcal{M}(C)_d \subseteq \mathcal{M}(C)_{d+1}$, one has $a_{\min,d} \le a_{\min,d+1}$. Furthermore, Theorem 5.5 implies that for each each $m \in \mathbb{N}$, there exists $d(m) \in \mathbb{N}$ such that $a - a_{\min}^\infty + \frac{1}{m} \in \mathcal{M}(C)_{d(m)}$. Thus one has

$$a_{\min}^\infty - \frac{1}{m} \le a_{\min,d(m)},$$

which implies that

$$\lim_{d \to \infty} a_{\min,d} = a_{\min}^\infty. \qquad \square$$

6.2. **Duality and state Hankel matrices.** Next, we introduce state Hankel and localizing matrices, which can be viewed as analogs of the noncommutative Hankel and localizing matrices (see e.g. [BKP16, Lemma 1.44]). By $\mathbf{W}_d^{\mathscr{S}}$ we denote the vector of all $\mathscr{S}$-words of degree at most d with respect to the degree lexicographic order; note that $\mathbf{W}_d^{\mathscr{S}}$ is of length $\Delta(n,d)$. Given $c \in \mathscr{S}$ denote $d_c := \lceil \frac{\deg c}{2} \rceil$. To $c \in \mathscr{S}$ and a linear functional $L : \mathscr{S}_{2d} \to \mathbb{R}$, we associate the following two matrices:

(a) the *Hankel matrix* $\mathbf{H}_d(L)$ is the symmetric matrix of size $\Delta(n,d)$, indexed by $\mathscr{S}$-words $u, v \in \mathscr{S}_d$, with $(\mathbf{H}_d(L))_{u,v} = L(\varsigma(u^\star v))$;
(b) the *localizing matrix* $\mathbf{H}_{d-d_c}(c\,L)$ is the symmetric matrix of size $\Delta(n, d-d_c)$, indexed by $\mathscr{S}$-words $u, v \in \mathscr{S}_{d-d_c}$, with $(\mathbf{H}_{d-d_c}(c\,L))_{u,v} = L(\varsigma(u^\star c v))$.

Note that the localizing matrix associated to L and 1 is simply the Hankel matrix associated to L.

Definition 6.2. A matrix M indexed by $\mathscr{S}$-words of degree $\leq d$ satisfies the *state Hankel condition* if and only if

$$M_{u,v} = M_{w,z} \text{ whenever } \varsigma(u^\star v) = \varsigma(w^\star z)\,. \tag{6.5}$$

Remark 6.3. Linear functionals on $\mathscr{S}_{2d}$ and matrices from $\mathrm{S}_{\Delta(n,d)}(\mathbb{R})$ satisfying the state condition (6.5) are in bijective correspondence. To a linear functional $L : \mathscr{S}_{2d} \to \mathbb{R}$, one can assign the matrix $\mathbf{H}_d(L)$, defined by $(\mathbf{H}_d(L))_{u,v} = L(\varsigma(u^\star v))$, satisfying the state Hankel condition, and vice versa.

The positivity of L relates to the positive semidefiniteness of its Hankel matrix $\mathbf{H}_d(L)$. The proof of the following lemma is straightforward and analogous to its free counterpart [BKP16, Lemma 1.44].

Lemma 6.4. *Given a linear functional $L : \mathscr{S}_{2d} \to \mathbb{R}$, one has $L(\varsigma(f^\star f)) \geq 0$ for all $f \in \mathscr{S}_d$, if and only if, $\mathbf{H}_d(L) \succeq 0$. Given $c \in \mathscr{S}$, one has $L(\varsigma(f^\star\, c\, f)) \geq 0$ for all $f \in \mathscr{S}_{d-d_c}$, if and only if, $\mathbf{H}_{d-d_c}(c\,L) \succeq 0$.*

We are now ready to state the dual SDP of (6.4).

Lemma 6.5. *The dual of (6.4) is the following SDP:*

$$\begin{aligned} \inf_{\substack{L:\mathscr{S}_{2d}\to\mathbb{R} \\ L \text{ linear}}} \quad & L(a) \\ \text{s.t.} \quad & (\mathbf{H}_d(L))_{u,v} = (\mathbf{H}_d(L))_{w,z}\,, \quad \textit{whenever } \varsigma(u^\star v) = \varsigma(w^\star z)\,, \\ & (\mathbf{H}_d(L))_{1,1} = 1\,, \\ & \mathbf{H}_{d-d_c}(c\,L) \succeq 0\,, \quad \textit{for all } c \in \{1\} \cup C \textit{ with } d_c \leq d\,. \end{aligned} \tag{6.6}$$

Proof. Let us denote by $(\mathcal{M}(C)_d)^\vee$ the dual cone of $\mathcal{M}(C)_d$. By (6.1), one has

$$(\mathcal{M}(C)_d)^\vee = \left\{ L : \mathscr{S}_{2d} \to \mathbb{R} \colon L \text{ linear}, L(f c f^\star) \geq 0 \text{ for all } c \in \{1\} \cup C,\ f \in \mathscr{S}_{d-d_c} \right\}.$$

By using a standard Lagrange duality approach, we obtain the dual of SDP (6.4):

$$a_{\min,d} = \sup_{a-m\in\mathcal{M}(C)_d} m = \sup_m \inf_{L\in(\mathcal{M}(C)_d)^\vee} (m + L(a-m)) \tag{6.7}$$

$$\leq \inf_{L\in(\mathcal{M}(C)_d)^\vee} \sup_m (m + L(a-m)) \tag{6.8}$$

$$= \inf_{L\in(\mathcal{M}(C)_d)^\vee} (L(a) + \sup_m m(1 - L(1))) \tag{6.9}$$

$$= \inf_L \{L(f) \colon L \in (\mathcal{M}(C)_d)^\vee,\ L(1) = 1\}, \tag{6.10}$$

The second equality in (6.7) comes from the fact that the inner minimization problem gives minimal value 0 if and only if $a - m \in \mathcal{M}(C)_d$. The inequality in (6.8) trivially holds. The inner maximization problem in (6.9) is bounded with maximum value 0 if and only $L(1) = 1$. Eventually, (6.10) is equivalent to SDP (6.6) by Remark 6.3 and Lemma 6.4. □

To establish strong duality for the SDP pair (6.4) and (6.6), we require a stronger version of algebraic boundedness.

Lemma 6.6. *Let $C \subseteq \mathscr{S}$ be balanced, and assume $N - x_1^2 - \cdots - x_n^2$ for some $N > 0$ is a conic combination of elements in $C \cap \mathbb{R}\langle \underline{x}\rangle_2$ and hermitian squares of elements in $\mathbb{R}\langle \underline{x}\rangle_1$. If $L \in (\mathcal{M}(C)_d)^\vee$ and $L(1) = 0$ then $L = 0$.*

Proof. For $k \in \mathbb{N}$ let K_k be the convex hull of

$$\left\{ f_0 f_0^\star, \varsigma(f_0 f_0^\star), f_1(N - x_j^2) f_1^\star, \varsigma(f_1(N - x_j^2) f_1^\star) \colon f_i \in \mathscr{S},\ \deg f_i \leq k - i,\ 1 \leq j \leq n \right\}.$$

We start by showing that for every $\mathscr{S}$-word w,

$$N^{\deg w} - w w^\star \in K_{\deg w}. \tag{6.11}$$

We proceed by induction on $\deg w$. Firstly, $N - x_j^2 \in K_1$, and then $N - \varsigma(x_j)^2 = \varsigma(N - x_j^2) + \varsigma((x_j - \varsigma(x_j))^2) \in K_1$. Now suppose (6.11) holds for $\mathscr{S}$-words of degree at most k. If w is a $\mathscr{S}$-word of degree $k+1$, there are three (partially overlapping) possibilities: $w = x_j v$ for a $\mathscr{S}$-word v of degree k, in which case

$$N^{\deg w} - w w^\star = N^{\deg v}(N - x_j^2) + x_j(N^{\deg v} - v v^\star) x_j \in K_{k+1};$$

or $w = \varsigma(x_j) v$ for a $\mathscr{S}$-word v of degree k, in which case

$$N^{\deg w} - w w^\star = N^{\deg v}(N - \varsigma(x_j)^2) + \varsigma(x_j)^2 (N^{\deg v} - v v^\star) \in K_{k+1};$$

or $w = \varsigma(x_j v)$ for a $\mathscr{S}$-word v of degree k, in which case

$$N^{\deg w} - w w^\star = N^{\deg w} - \varsigma((x_j v)(x_j v)^\star) + \varsigma((x_j v - \varsigma(x_j v))(x_j v - \varsigma(x_j v))^\star) \in K_{k+1}.$$

Now assume $L \in (\mathcal{M}(C)_d)^\vee$ and $L(1) = 0$. Since $\varsigma(K_{2d}) \subseteq \mathcal{M}(C)_d$ by the assumption on C, we have $L(\varsigma(w w^\star)) = 0$ for all $\mathscr{S}$-words w of degree at most d by (6.11). Then $\varsigma((u \pm v^\star)(u \pm v^\star)^\star) \in \mathcal{M}(C)_d$ implies $L(\varsigma(uv)) = 0$ for all $\mathscr{S}$-words u, v of degree at most d. Since $\mathscr{S}_{2d}$ is spanned by such $\varsigma(uv)$, we conclude $L = 0$. □

Proposition 6.7. *Let $C \subseteq \mathscr{S}$ be balanced, and assume $N - x_1^2 - \cdots - x_n^2$ for some $N > 0$ is a conic combination of elements in $C \cap \mathbb{R}\langle \underline{x} \rangle_2$ and hermitian squares of elements in $\mathbb{R}\langle \underline{x} \rangle_1$; e.g., $N - x_1^2 - \cdots - x_n^2 \in C$. Then SDP (6.4) satisfies strong duality, i.e., there is no duality gap between SDP (6.6) and SDP (6.4).*

Proof. Suppose SDP (6.4) is feasible. Then $a - a_{\min,d}$ is a boundary point of the cone $\mathcal{M}(C)_d$ in $\mathscr{S}_{2d}$. Therefore there is a supporting hyperplane for $\mathcal{M}(C)_d$ through $a - a_{\min,d}$. In other words, there is a nonzero linear functional $L \in (\mathcal{M}(C)_d)^\vee$ such that $L(a - a_{\min,d}) = 0$. By Lemma 6.6 we have $L(1) > 0$. After rescaling we can then assume that $L(1) = 1$, and so $L(a - a_{\min,d}) = 0$ implies $L(a) = a_{\min,d}$. Therefore there is no duality gap. □

Remark 6.8. The condition on the constraint set C in Proposition 6.7 is stronger that algebraic boundedness. Nevertheless, it is satisfied in many prominent instances, for example if C contains $\pm(x_j^2 - 1)$ for all j (optimization over binary observables) as in Section 7 below, or if C contains $\pm(x_j^2 - x_j)$ for all j (optimization over projections) as in Section 8. Furthermore, if C is algebraically bounded, or $\mathcal{D}_C^\infty$ is known to be bounded in operator norm, then we can simply add the desired constraint directly to C, thus fulfilling the assumption of Proposition 6.7 without affecting $\mathcal{D}_C^\infty$.

6.3. **Minimizer extraction.** The goal of this subsection is to derive an algorithm to extract minimizers and certify exactness of state polynomial optimization problems. The forthcoming statements can be seen as state variants of the results derived in the context of commutative polynomials [CF98], eigenvalue optimization of noncommutative polynomials [PNA10, BKP16], and optimization of trace polynomials [KMV22, BCKP13].

Definition 6.9. Suppose $L : \mathscr{S}_{2d+2\delta} \to \mathbb{R}$ is a linear functional with restriction $\tilde{L} : \mathscr{S}_{2d} \to \mathbb{R}$. We associate to L and $\tilde{L}$ the Hankel matrices $\mathbf{H}_{d+\delta}(L)$ and $\mathbf{H}_d(\tilde{L})$ respectively, and get the block form

$$\mathbf{H}_{d+\delta}(L) = \begin{bmatrix} \mathbf{H}_d(\tilde{L}) & B \\ B^T & B' \end{bmatrix}.$$

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

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

$$\mathcal{K} := \operatorname{span}\{\mathbf{u} \colon \deg u \le d + \delta\} = \operatorname{span}\{\mathbf{u} \colon \deg u \le d\},$$

where the equality is a consequence of the flatness assumption. Afterwards, one can consider, for each $p \in \mathscr{S}$, the multiplication operator X_p on $\mathcal{K}$ and the $\star$-representation $\pi : \mathscr{S} \to \mathcal{B}(\mathcal{K})$ defined by $\pi(p) = X_p$. Let $\mathbf{v}$ be the vector representing 1 in $\mathcal{K}$; then $L(p) = \langle \pi(p)\mathbf{v}, \mathbf{v} \rangle$ for all $p \in \mathscr{S}$. In general, elements of $\pi(\mathscr{S})$ are central in $\pi(\mathscr{S})$; if they are actually scalar multiples of the identity on $\mathcal{K}$, then π is not just a $\star$-representation, but it respects the state symbol in the sense that $\pi(f) = f(\pi(x_1), \dots, \pi(x_n))$ for every $f \in \boldsymbol{\mathscr{S}}$. This fact applies to our SDP hierarchy as follows.

Proposition 6.10. *Let $a \in \mathscr{S}$, suppose that $C \subseteq \boldsymbol{\mathscr{S}}$ satisfies the assumptions of Proposition 6.7, and let $d, \delta \in \mathbb{N}$ be such that $d + \delta \geq d_a, d_c$ for $c \in C$. Assume that L is a δ-flat optimal solution of SDP (6.6), and let the $\star$-representation $\pi : \boldsymbol{\mathscr{S}} \to \mathcal{B}(\mathcal{K})$ and the unit vector $\mathbf{v} \in \mathcal{K}$ be constructed as above. If $\pi(\mathscr{S}) = \mathbb{R}$, then*

(i) $(\lambda, \underline{X}) \in \vec{\mathcal{D}}_C$ *where* $\underline{X} = (\pi(x_1), \dots, \pi(x_n))$ *and* $\lambda(Y) = \langle Y\mathbf{v}, \mathbf{v}\rangle$;
(ii) $L(p) = p(\lambda; \underline{X})$ *for all* $p \in \mathscr{S}$;
(iii) $a_{\min,d+\delta} = L(a) = a^{\infty}_{\min}$.

Proof. As seen in the paragraph before Proposition 6.10, the operators X_i are well-defined (thanks to the flatness assumption) and symmetric. After choosing an orthonormal basis of $\mathcal{K}$ we can view X_i as $r \times r$ symmetric matrices. Moreover, $L(p) = \lambda(p)$ for all $p \in \mathscr{S}$. Since $\pi(\mathscr{S}) = \mathbb{R}$, we furthermore have $L(p) = p(\lambda; \underline{X})$ for all $p \in \mathscr{S}$, so (ii) holds. For $c \in C$, one in particular has $c(\lambda; \underline{X}) = L(c) \geq 0$ because $\mathbf{H}_{d-d_c}(c\,L) \succeq 0$ as L is a feasible solution of SDP (6.6). Hence (i) holds. Proposition 6.7 implies $a_{\min,d+\delta} = L(a) \leq a^{\infty}_{\min}$; on the other hand, $a^{\infty}_{\min} \leq a(\lambda; \underline{X}) = L(a)$, and therefore (iii) holds. □

Remark 6.11. The condition $\pi(\mathscr{S}) = \mathbb{R}$ in Proposition 6.10 in particular holds if L is an extreme optimal solution of (6.6). 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 minimum rank, yielding an extreme linear functional; see [LLR08, §4.4.1] for more details.

6.4. **Reduction by exploiting sparsity.** In this subsection, we briefly introduce the approaches for reducing the sizes of SDP (6.6) and SDP (6.4) by exploiting sparsity encoded in the state polynomial optimization problem, which are adapted from the case of eigenvalue and trace optimization over nc polynomials [KMP22, WM21].

6.4.1. *Correlative sparsity.* For $I \subseteq [n] := \{1, \dots, n\}$, let $\mathscr{S}_I \subseteq \mathscr{S}$ (resp. $\boldsymbol{\mathscr{S}}_I \subseteq \boldsymbol{\mathscr{S}}$) be the subset of state (resp. nc state) polynomials in variables $x_i, i \in I$ only. Let $I_1, \dots, I_\ell \subseteq [n]$ be a tuple of index sets and further $J_1, \dots, J_\ell \subseteq C$ be a partition of the constraint polynomials in C such that

$$a \in \mathscr{S}_{I_1} + \cdots + \mathscr{S}_{I_\ell}; \tag{6.12}$$

$$J_k \subseteq \boldsymbol{\mathscr{S}}_{I_k} \text{ for } k = 1, \dots, \ell. \tag{6.13}$$

The tuple of index sets $I_1, \dots, I_\ell$ is then called the *correlative sparsity pattern* of (6.2) and (6.3). We build the Hankel submatrix $\mathbf{H}^{I_k}_d(L)$ (resp. the localizing submatrix $\mathbf{H}^{I_k}_{d-d_c}(c\,L)$) with respect to the correlative sparsity pattern by retaining only rows and columns indexed by $\mathbf{W}^{\boldsymbol{\mathscr{S}}_{I_k}}_d$ (resp. $\mathbf{W}^{\boldsymbol{\mathscr{S}}_{I_k}}_{d-d_c}$) for each $k \in [\ell]$ (resp. each $c \in J_k$).

Let us consider the correlative sparsity adapted version of (6.6):

$$
\begin{aligned}
\inf_{\substack{L:\mathscr{S}_{2d}\to\mathbb{R}\\ L\ linear}} \quad & L(a) \\
\text{s.t.} \quad & (\mathbf{H}_d^{I_k}(L))_{u,v} = (\mathbf{H}_d^{I_k}(L))_{w,z}\,, \quad \text{whenever } \varsigma(u^\star v) = \varsigma(w^\star z),\ \text{for } k\in[\ell]\,, \\
& (\mathbf{H}_d(L))_{1,1} = 1\,, \\
& \mathbf{H}_{d-d_c}^{I_k}(c\,L) \succeq 0\,, \quad \text{for all } c\in\{1\}\cup J_k \text{ with } d_c\le d \text{ and } k\in[\ell]\,,
\end{aligned}
\tag{6.14}
$$

with optimum denoted by $a^{\mathrm{cs}}_{\min,d}$.

Theorem 6.12. *Let $C\subseteq\boldsymbol{\mathscr{S}}$ be balanced and algebraically bounded, and let $a\in\mathscr{S}$. Suppose that there exist subsets $I_1,\dots,I_\ell$ and $J_1,\dots,J_\ell$ such that* (6.12) *and* (6.13) *hold, and that $I_1,\dots,I_\ell$ satisfy the running intersection property (RIP), i.e., for every $k\in[\ell-1]$, we have that*

$$
\left(I_{k+1}\cap\bigcup_{j\le k} I_j\right)\subseteq I_i \quad \textit{for some } i\le k. \tag{6.15}
$$

Then $\lim_{d\to\infty} a^{\mathrm{cs}}_{\min,d} = a^{\infty}_{\min}$.

Proof. Let us define

$$
\mathrm{QM}(C^\varsigma)^{\mathrm{cs}} := \mathrm{QM}(J_1^\varsigma)+\cdots+\mathrm{QM}(J_\ell^\varsigma).
$$

To obtain the convergence result, we need to prove a sparse analog of Theorem 5.5, namely $a(\lambda;\underline{X})\ge 0$ for all $(\lambda;\underline{X})\in\vec{\mathcal{D}}^\infty_C$ implies $a+\varepsilon\in\mathrm{QM}(C^\varsigma)^{\mathrm{cs}}$, for all $\varepsilon>0$. This sparse representation result requires an adaptation of [KMP22, Theorem 3.3], so we only sketch the main steps while emphasizing changes required. The proof is by contradiction, so we suppose that $a+\varepsilon\notin\mathrm{QM}(C^\varsigma)^{\mathrm{cs}}$ for some $\varepsilon>0$. By the algebraically bounded assumption, each quadratic module $\mathrm{QM}(J_k^\varsigma)$ in $\mathscr{S}_{I_k}$ is archimedean (see Lemma 5.4). In particular, 1 is an algebraic interior point of $\mathrm{QM}(C^\varsigma)^{\mathrm{cs}}$ in $\mathscr{S}_{I_1}+\cdots+\mathscr{S}_{I_\ell}$. Thus by the Eidelheit-Kakutani separation theorem (a version of the Hahn-Banach separation theorem suitable for this context; see [Bar02, Corollary III.1.7]), there exists a unital linear functional $\varphi:\mathscr{S}_{I_1}+\cdots+\mathscr{S}_{I_\ell}\to\mathbb{R}$ with $\varphi(\mathrm{QM}(C^\varsigma)^{\mathrm{cs}})\subseteq\mathbb{R}_{\ge0}$ and $\varphi(a)<0$. We pick any extension of φ to a linear functional on $\mathscr{S}$ which we again denote by φ. Let $L:\mathbb{R}\langle\underline{x}\rangle\to\mathbb{R}$ be the unital $\star$-functional given by $L(p):=\varphi(\varsigma(p))$, and let us denote by L^k the restriction of L to $\mathbb{R}\langle\underline{x}(I_k)\rangle$.

Then we proceed exactly as in the proof of Theorem 5.5 for each $k\in\{1,\dots,\ell\}$. Since each quadratic module $\mathrm{QM}(J_k^\varsigma)\subseteq\mathscr{S}_{I_k}$ is archimedean, one can apply Proposition 5.1 to obtain a vector state $\lambda_k\in\mathcal{S}(\mathcal{H}_k)$ and a tuple of self-adjoint operators $\underline{X}^k\in\mathcal{B}(\mathcal{H}_k)^{|I_k|}$ such that $L^k(p)=\lambda_k(p(\underline{X}^k))$ for all $p\in\mathbb{R}\langle\underline{x}(I_k)\rangle$, and $(\lambda_k,\underline{X}^k)\in\vec{\mathcal{D}}^\infty_{J_k}$. Here $\mathcal{H}_k$ denotes the Hilbert space completion of the quotient of $\mathbb{R}\langle\underline{x}(I_k)\rangle$ by the set of nullvectors corresponding to L^k, obtained through the GNS construction. Now, the proof proceeds by induction on ℓ

to show that there exist a vector state λ and a tuple of self-adjoint operators $\underline{X}$ such that $b(\lambda, \underline{X}) = \varphi(b)$ for all $b \in \mathscr{S}_{I_1} + \cdots + \mathscr{S}_{I_\ell}$, and $(\lambda, \underline{X}) \in \vec{\mathcal{D}}_C^\infty$.

We focus specifically on the case $\ell = 2$, as the general case then follows by an inductive argument relying on the running intersection property, similarly to the proof of [KMP22, Theorem 3.3]. We denote by L^{12} the restriction of L^1 to $\mathbb{R}\langle \underline{x}(I_1 \cap I_2)\rangle$ and again we apply Proposition 5.1 to obtain a vector state $\lambda_{12} \in \mathcal{S}(\mathcal{H}_{12})$ and self-adjoint $\underline{X}^{12} \in \mathcal{B}(\mathcal{H}_{12})^{|I_1 \cap I_2|}$ such that $L^{12}(p) = \lambda_{12}(p(\underline{X}^{12}))$, for all $p \in \mathbb{R}\langle \underline{x}(I_1 \cap I_2)\rangle$. For $k \in \{1, 2\}$, we denote by i_k the canonical embedding from $\mathbb{R}\langle \underline{x}(I_1 \cap I_2)\rangle$ to $\mathbb{R}\langle \underline{x}(I_k)\rangle$. Let ι_k be the canonical embedding from $\mathcal{B}(\mathcal{H}_{12})$ to $\mathcal{B}(\mathcal{H}_k)$, satisfying $\iota_k(X_i^{12}) = X_i^k$ for all $i \in I_1 \cap I_2$. Then we apply [KMP22, Theorem 3.1] to obtain an amalgamation $\mathcal{A}$ with state λ and homomorphisms $j_k : \mathcal{B}(\mathcal{H}_k) \to \mathcal{A}$ such that $j_1 \circ \iota_1 = j_2 \circ \iota_2$. After performing the GNS construction with $(\mathcal{A}, \lambda)$, we obtain a Hilbert space $\mathcal{K}$, a representation $\pi : \mathcal{A} \to \mathcal{B}(\mathcal{K})$ and a vector $\xi \in \mathcal{K}$ so that $\lambda(b) = \langle \pi(b)\xi, \xi\rangle$. We next define $\underline{X} := (X_1, \dots, X_n)$, with $X_i := \pi(j_1(X_i^1))$ if $i \in I_1$ and $X_i := \pi(j_2(X_i^2))$ otherwise. This tuple of operators is well-defined thanks to the amalgamation property.

Let us now set $\tilde{L}(p) := \langle p(\underline{X})\xi, \xi\rangle$, for all $p \in \mathbb{R}\langle \underline{x}\rangle$. We prove that $\tilde{L}$ agrees with L^k (as well as L) on $\mathbb{R}\langle \underline{x}(I_k)\rangle$ thanks to the amalgamation setup. Indeed, for all $p \in \mathbb{R}\langle \underline{x}(I_k)\rangle$ one has

$$\begin{aligned}\tilde{L}(p) &= \langle p(\underline{X})\xi, \xi\rangle = \langle p(\pi(j_k(\underline{X}^k)))\xi, \xi\rangle = \langle \pi(p(j_k(\underline{X}^k)))\xi, \xi\rangle \\ &= \lambda(p(j_k(\underline{X}^k))) = \lambda(j_k(p(\underline{X}^k))) = \lambda_k(p(\underline{X}^k)) = L^k(p) = L(p).\end{aligned}$$

Hence, this yields $\varphi(b) = b(\lambda, \underline{X})$ for all $b \in \mathscr{S}_k$ and by linearity of φ, we obtain $\varphi(b) = b(\lambda, \underline{X})$ for all $b \in \mathscr{S}_1 + \mathscr{S}_2$. To show that $(\lambda, \underline{X}) \in \vec{\mathcal{D}}_C^\infty$, we proceed as in the proof of Theorem 5.5. □

6.4.2. *Sign symmetry.* For $a \in \boldsymbol{\mathscr{S}}$ and a binary vector $s \in \{0, 1\}^n$, let $[a]_s \in \boldsymbol{\mathscr{S}}$ be defined by $[a]_s(x_1, \dots, x_n) := a((-1)^{s_1}x_1, \dots, (-1)^{s_n}x_n)$. Then a is said to have the sign symmetry represented by a binary vector $s \in \{0, 1\}^n$ if $[a]_s = a$. We use $S(a) \subseteq \{0, 1\}^n$ to denote all sign symmetries of a and let $S(C) := \cap_{c \in C} S(c)$ for $C \subseteq \boldsymbol{\mathscr{S}}$. Denote

$$\mathcal{U} = \{u \in \mathbf{W}_{2d}^{\boldsymbol{\mathscr{S}}} \colon S(\{a\} \cup C) \subseteq S(u)\},$$

and let $\mathscr{S}_{\mathcal{U}} \subseteq \mathscr{S}_{2d}$ be the span of $\{\varsigma(u) \colon u \in \mathcal{U}\}$. Consider the optimization problem given by (6.3). We can build a block-diagonal SDP hierarchy for (6.3) by exploiting its sign symmetries. To this end, we define an equivalence relation $\sim$ on $\mathbf{W}_d^{\boldsymbol{\mathscr{S}}}$ by

$$u \sim v \iff u^\star v \in \mathcal{U}. \tag{6.16}$$

The equivalence relation $\sim$ gives rise to a partition of $\mathbf{W}_d^{\boldsymbol{\mathscr{S}}}$:

$$\mathbf{W}_d^{\boldsymbol{\mathscr{S}}} = \bigsqcup_{i=1}^{p_d} \mathbf{W}_{d,i}^{\boldsymbol{\mathscr{S}}}. \tag{6.17}$$

We build the Hankel submatrix $\mathbf{H}_{d,i}(L)$ (resp. the localizing submatrix $\mathbf{H}_{d-d_c,i}(c\,L)$) with respect to the sign symmetry by retaining only those rows and columns that are indexed by $\mathbf{W}^{\mathscr{S}}_{d,i}$ (resp. $\mathbf{W}^{\mathscr{S}}_{d-d_c,i}$) for each $i \in [p_d]$ (resp. $i \in [p_{d_c}]$).

Let us consider the sign symmetry adapted version of (6.6):

$$
\begin{aligned}
\inf_{\substack{L:\mathscr{S}_{\mathcal{U}}\to\mathbb{R}\\ L \text{ linear}}} \quad & L(a) \\
\text{s.t.} \quad & (\mathbf{H}_{d,i}(L))_{u,v} = (\mathbf{H}_{d,i}(L))_{w,z}\,, \quad \text{whenever } \varsigma(u^\star v) = \varsigma(w^\star z), \text{ for } i \in [p_d]\,, \\
& L(1) = 1\,, \\
& \mathbf{H}_{d-d_c,i}(c\,L) \succeq 0\,, \quad \text{for all } c \in \{1\} \cup C \text{ with } d_c \le d \text{ and } i \in [p_{d_c}]\,,
\end{aligned}
\tag{6.18}
$$

with optimum denoted by $a^{\mathrm{ss}}_{\min,d}$.

Theorem 6.13. *We have that* $a^{\mathrm{ss}}_{\min,d} = a_{\min,d}$.

Proof. For a linear functional $L : \mathscr{S}_{2d} \to \mathbb{R}$ and $s \in \{0,1\}^n$, let $L^s : \mathscr{S}_{2d} \to \mathbb{R}$ be another linear functional given by $L^s(u) = L([u]_s)$. Suppose that L is an optimal solution of (6.6) and let $L' = \frac{1}{|S(\{a\}\cup C)|}\sum_{s\in S(\{a\}\cup C)} L^s$ which is also an optimal solution of (6.6). We claim that $L'(\varsigma(u^\star v)) = 0$ whenever $u \not\sim v \in \mathbf{W}^{\mathscr{S}}_d$. By (6.16), if $u \not\sim v$, then there exists $s' \in S(\{a\}\cup C)$ such that $[u^\star v]_{s'} = -u^\star v$. We then have

$$
\begin{aligned}
L'(\varsigma(u^\star v)) &= \frac{1}{|S(\{a\}\cup C)|} \sum_{s\in S(\{a\}\cup C)} L^s(\varsigma(u^\star v)) = -\frac{1}{|S(\{a\}\cup C)|} \sum_{s\in S(\{a\}\cup C)} L^s(\varsigma([u^\star v]_{s'})) \\
&= -\frac{1}{|S(\{a\}\cup C)|} \sum_{s\in S(\{a\}\cup C)} L^{s+s'}(\varsigma(u^\star v)) = -\frac{1}{|S(\{a\}\cup C)|} \sum_{s\in S(\{a\}\cup C)} L^s(\varsigma(u^\star v)) \\
&= -L'(\varsigma(u^\star v)).
\end{aligned}
$$

Thus $L'(\varsigma(u^\star v)) = 0$ as desired. From this we see that the restriction of L' to $\mathscr{S}_{\mathcal{U}}$ is a feasible solution of (6.18) and so $a^{\mathrm{ss}}_{\min,d} \le a_{\min,d}$.

On the other hand, let $L : \mathscr{S}_{\mathcal{U}} \to \mathbb{R}$ be an optimal solution of (6.18). We define a functional $L' : \mathscr{S}_{2d} \to \mathbb{R}$ as follows:

$$
L'(\varsigma(u)) = \begin{cases} L(\varsigma(u)), & \text{if } u \in \mathcal{U}, \\ 0, & \text{otherwise.} \end{cases}
$$

One can easily check that L' is a feasible solution of (6.6). So $a^{\mathrm{ss}}_{\min,d} \ge a_{\min,d}$ and it follows $a^{\mathrm{ss}}_{\min,d} = a_{\min,d}$ as desired. □

Remark 6.14. In nc polynomial optimization, the exploitation of sign symmetries can be extended to an iterative procedure via the more general notion of "term sparsity" [WM21]. Due to the multitude of technical details and overhead involved, the state polynomial version of term sparsity will be explored elsewhere in the future.

7. Nonlinear Bell inequalities

In this section we connect state polynomial optimization to violations of nonlinear Bell inequalities, establish a further reduction of our optimization procedures based on conditional expectation that is tailored to the quantum-mechanical formalism (Proposition 7.1), and outline a few examples.

For the sake of simplicity, we restrict to bipartite models where two parties share a state and use binary observables to produce measurements. A (m-input 2-output) *quantum commuting model* is then given as a triple $(\lambda, \underline{A}, \underline{B})$ where $\lambda \in \mathcal{S}(\mathcal{H})$ is a state and $\underline{A} = (A_1, \dots, A_m), \underline{B} = (B_1, \dots, B_m)$ are commuting tuples of binary observables in $\mathcal{B}(\mathcal{H})$:

$$A_i^* = A_i, \quad A_i^2 = I, \quad B_j^* = B_j, \quad B_j^2 = I, \quad A_iB_j = B_jA_i$$

for all $1 \leq i, j \leq m$. The correlations produced by $(\lambda, \underline{A}, \underline{B})$ are determined by $\lambda(A_iB_j)$ for $i, j = 0, \dots m$ where $A_0 = B_0 = I$. If $\mathcal{H} = \mathcal{H}' \otimes \mathcal{H}'$ for a finite-dimensional $\mathcal{H}'$ and $A_i = A_i' \otimes I$, $B_j = I \otimes B_j'$ then $(\lambda, \underline{A}, \underline{B})$ is a (finite-dimensional) *spatial quantum model.* In this case, $\dim \mathcal{H}'$ is the local dimension of $(\lambda, \underline{A}, \underline{B})$, and λ is usually given by a density matrix. On the other hand, if $\underline{A}$ and $\underline{B}$ are tuples of commuting operators, then $(\lambda, \underline{A}, \underline{B})$ is *classical.* In this case, the correlations can be obtained as expectations of products of binary random variables on a probability space.

To warm up, consider the expression

$$\lambda(A_1B_1) + \lambda(A_1B_2) + \lambda(A_2B_1) - \lambda(A_2B_2) \tag{7.1}$$

for a model $(\lambda, \underline{A}, \underline{B})$. The classical Bell inequality states that (7.1) is at most 2 for classical models. On the other hand, (7.1) attains the value $2\sqrt{2}$ for a spatial quantum model with local dimension 2. Furthermore, Tsirelson's bound implies that the value $2\sqrt{2}$ is optimal for all quantum commuting models. From the perspective of this paper, Tsirelson's bound can be recovered as a state polynomial optimization problem

$$\begin{aligned} \sup\ & \varsigma(x_1y_1) + \varsigma(x_1y_2) + \varsigma(x_2y_1) - \varsigma(x_2y_2) \\ \text{s.t.}\ & x_i^2 = 1,\ y_j^2 = 1,\ [x_i, y_j] = 0. \end{aligned}$$

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 (7.1) by eigenvalue-optimizing $x_1y_1 + x_1y_2 + x_2y_1 - x_2y_2$ subject to $x_j^2 = y_j^2 = 1$ and $[x_i, y_j] = 0$. On the other hand, 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. Even for classical models, nonlinearity complicates the study of polynomial Bell inequalities; for example, the supremum of a Bell-like expression over classical models can be strictly larger than the supremum over deterministic models. Nonlinearity also renders noncommutative polynomial eigenvalue optimization, which is commonly used to bound quantum violations of linear Bell inequalities, inapplicable to polynomial Bell inequalities. On the other hand, state polynomial optimization gives upper bounds on violations of polynomial Bell inequalities.

7.1. **Universal algebras of binary observables.** In this section we derive further simplifications for optimization of a state polynomial subject to a balanced constraint set of the form

$$C = \left\{\pm(1-x_1^2), \cdots, \pm(1-x_n^2), \pm[x_{i_1}, x_{j_1}], \dots, \pm[x_{i_\ell}, x_{j_\ell}]\right\} \tag{7.2}$$

for some $1 \le i_k, j_k \le n$. As mentioned above, optimal Bell inequalities correspond to optimization problems subject to constraint sets of the form (7.2). By Corollary 6.1, every state polynomial optimization problem on (7.2) admits a convergent SDP hierarchy as in Section 6.1, and these SDPs satisfy strong duality by Proposition 6.7.

Let G be a group. Analogously to the construction of nc state polynomials, one can define the *state group algebra* $\boldsymbol{\mathscr{S}}(G)$ of G: namely, let $\mathscr{S}(G)$ be the real polynomial ring in commutative symbols $\varsigma(g)$ for $g \in G \setminus \{1\}$, subject to $\varsigma(g^{-1}) = \varsigma(g)$, and let $\boldsymbol{\mathscr{S}}(G) = \mathscr{S}(G) \otimes \mathbb{R}[G]$, where $\mathbb{R}[G]$ is the real group $\star$-algebra of G, where the involution is given by $g^\star = g^{-1}$ for $g \in G$. As before, there is a natural map $\boldsymbol{\mathscr{S}}(G) \to \mathscr{S}(G)$. Let $\Sigma\boldsymbol{\mathscr{S}}(G)^2$ denote the set of sums of hermitian squares $ff^\star$ for $f \in \boldsymbol{\mathscr{S}}(G)$.

Returning back to (7.2), consider the group

$$G = \langle x_1, \dots, x_n \mid x_1^2 = \cdots = x_n^2 = 1, x_{i_1}x_{j_1} = x_{j_1}x_{i_1}, \dots, x_{i_\ell}x_{j_\ell} = x_{j_\ell}x_{i_\ell}\rangle.$$

Let $\pi : \langle \underline{x}\rangle \to G$ be the canonical homomorphism. We extend it to a ς-respecting $\star$-homomorphism $\pi : \boldsymbol{\mathscr{S}} \to \boldsymbol{\mathscr{S}}(G)$. Then for $a \in \mathscr{S}$,

$$a \in \mathrm{QM}(C^\varsigma) \qquad \Longleftrightarrow \qquad \pi(a) \in \varsigma\left(\Sigma\boldsymbol{\mathscr{S}}(G)^2\right). \tag{7.3}$$

The relation (7.3) is advantageous in optimizing state polynomials subject to C: the sizes of SDPs (6.4) and (6.6) can be reduced by indexing with a basis of $\boldsymbol{\mathscr{S}}(G)$, and only a single semidefinite constraint is needed (corresponding to $\Sigma\boldsymbol{\mathscr{S}}(G)^2$). This reduction is used in all subsequent computational examples.

A further reduction is sometimes possible. Given $f \in \mathscr{S}(G)$, its *support* are elements of G appearing in f.

Proposition 7.1. *Let $a \in \mathscr{S}$, and let $H \subseteq G$ be the subgroup generated by the support of $\pi(a)$. Then*

$$a \in \mathrm{QM}(C^\varsigma) \qquad \Longleftrightarrow \qquad \pi(a) \in \varsigma\left(\Sigma\boldsymbol{\mathscr{S}}(H)^2\right). \tag{7.4}$$

Proof. Let $\mathbf{1}_H : G \to H \cup \{0\}$ be the indicator function, where $\mathbf{1}_H(g) = g$ if $g \in H$ and $\mathbf{1}_H(g) = 0$ otherwise. Let $E : \boldsymbol{\mathscr{S}}(G) \to \boldsymbol{\mathscr{S}}(H)$ be the unital $\star$-linear map given by

$$E\big(\varsigma(g_1)\cdots\varsigma(g_\ell)g_0\big) = \varsigma(\mathbf{1}_H(g_1))\cdots\varsigma(\mathbf{1}_H(g_\ell))\mathbf{1}_H(g_0).$$

Note that E commutes with ς, restricts to a ring homomorphism $\mathscr{S}(G) \to \mathscr{S}(H)$, and has conditional expectation properties (cf. [SS13, Section 3]):

(i) $E(b_1ab_2) = b_1E(a)b_2$ for $a \in \boldsymbol{\mathscr{S}}(G)$ and $b_1, b_2 \in \boldsymbol{\mathscr{S}}(H)$,

(ii) $E(\Sigma\boldsymbol{\mathscr{S}}(G)^2) = \Sigma\boldsymbol{\mathscr{S}}(H)^2$.

The second property follows by [SS13, Proposition 3.4] and $E : \mathscr{S}(G) \to \mathscr{S}(H)$ being a homomorphism. Since $\pi(a) \in \mathscr{S}(H)$, (ii) implies $\pi(a) \in \varsigma(\Sigma\boldsymbol{\mathscr{S}}(H)^2)$ if and only if $\pi(a) \in \varsigma(\Sigma\boldsymbol{\mathscr{S}}(G)^2)$, and the rest follows by (7.3). □

If H is a proper subgroup of G, the sizes of SDPs (6.4) and (6.6) can thus be further decreased by Proposition 7.1; this is illustrated in Example 7.2.1 below.

7.2. **Examples.** We demonstrate the optimization results from Section 6 on the following polynomial Bell inequalities. The codes for reproducing the results are available at

`https://github.com/wangjie212/NCTSSOS/blob/master/examples/stateopt.jl`

For all examples, we employ `MOSEK 10.0` as an SDP solver. For more details on the modeling syntax, we refer the interested programmer to the tutorial from [MW23, Appendix B.2] that describes a similar syntax to perform trace polynomial optimization.

7.2.1. *Example.* One of the first considered polynomial Bell inequalities is

$$\lambda(A_1B_2 + A_2B_1)^2 + \lambda(A_1B_1 - A_2B_2)^2 \leq 4 \tag{7.5}$$

given in [Uff02], where it is shown that (7.5) holds for all classical models, and for all spatial quantum models with local dimension 2 (the equality is obtained for a model with a maximally entangled state). In [NKI02], (7.5) is shown to hold for all spatial quantum models. An automatized proof of (7.5) for arbitrary quantum commuting models can be obtained by solving the optimization problem

$$\begin{aligned} &\sup\ \left(\varsigma(x_1y_2) + \varsigma(x_2y_1)\right)^2 + \left(\varsigma(x_1y_1) - \varsigma(x_2y_2)\right)^2 \\ &\quad\text{s.t.}\ \ x_i^2 = 1, y_j^2 = 1,\ [x_i, y_j] = 0 \text{ for } i,j = 1,2. \end{aligned} \tag{7.6}$$

Let $C = \{\pm(1 - x_j^2), \pm(1 - y_j^2), \pm[x_i, y_j]\}_{i,j=1,2}$. The relaxation of (7.6) with $d = 3$ as in Section 6.1,

$$\inf\ \mu\ \text{ s.t. }\ \mu - \left(\varsigma(x_1y_2) + \varsigma(x_2y_1)\right)^2 - \left(\varsigma(x_1y_1) - \varsigma(x_2y_2)\right)^2 \in \mathcal{M}(C)_d \tag{7.7}$$

outputs 4, which coincides with the classical value in (7.5). The concrete implementation of (7.7) encodes the relations $x_j^2 = y_j^2 = 1$ and $[x_i, y_j] = 0$ directly into the SDP, as in Section 7.1. The resulting SDP has 2032 variables, and a 209×209 semidefinite constraint.

Alternatively, we can also invoke Proposition 7.1. The support of $\left(\varsigma(x_1y_2) + \varsigma(x_2y_1)\right)^2 - \left(\varsigma(x_1y_1) - \varsigma(x_2y_2)\right)^2$ in

$$G = \langle x_i, y_j \mid x_i^2 = y_j^2 = 1, x_iy_j = y_jx_i \text{ for } i,j = 1,2\rangle$$

is $\{x_iy_j\}_{i,j=1,2}$, which generates the subgroup H of G consisting of all words in generators x_i, y_j of even length. Cutting down the aforementioned SDP with respect to H then results in an SDP with 933 variables and a 112×112 semidefinite constraint, which returns the value 4 in shorter time.

7.2.2. *Example.* Polynomial Bell inequalities also arise from covariances of quantum correlations. Let

$$\mathrm{cov}_\lambda(X,Y)=\lambda(XY)-\lambda(X)\lambda(Y).$$

In [PHBB17] it is shown that while

$$\begin{aligned}&\mathrm{cov}_\lambda(A_1,B_1)+\mathrm{cov}_\lambda(A_1,B_2)+\mathrm{cov}_\lambda(A_1,B_3)\\ &+\mathrm{cov}_\lambda(A_2,B_1)+\mathrm{cov}_\lambda(A_2,B_2)-\mathrm{cov}_\lambda(A_2,B_3)\\ &+\mathrm{cov}_\lambda(A_3,B_1)-\mathrm{cov}_\lambda(A_3,B_2)\end{aligned}\tag{7.8}$$

is at most $\frac{9}{2}$ for classical models, it attains the value 5 for a spatial quantum model of local dimension 2 and a maximally entangled state. The authors also performed extensive numerical search within spatial quantum models with local dimension at most 5, but no higher value of (7.8) was found. They left it as an open question whether higher dimensional entangled states could lead to larger violations [PHBB17, Appendix D.1(b)].

Let

$$\begin{aligned}b=&\varsigma(x_1y_1)-\varsigma(x_1)\varsigma(y_1)+\varsigma(x_1y_2)-\varsigma(x_1)\varsigma(y_2)+\varsigma(x_1y_3)-\varsigma(x_1)\varsigma(y_3)\\ &+\varsigma(x_2y_1)-\varsigma(x_2)\varsigma(y_1)+\varsigma(x_2y_2)-\varsigma(x_2)\varsigma(y_2)-\varsigma(x_2y_3)+\varsigma(x_2)\varsigma(y_3)\\ &+\varsigma(x_3y_1)-\varsigma(x_3)\varsigma(y_1)-\varsigma(x_3y_2)+\varsigma(x_3)\varsigma(y_2)\,.\end{aligned}$$

The relaxation of

$$\sup\ b\ \text{ s.t. }\ x_i^2=1,\, y_j^2=1,\, [x_i,y_j]=0 \text{ for } i,j=1,2,3$$

with $d=2$ returns 5. Therefore the value of (7.8) is at most 5 for all quantum commuting models.

7.2.3. *Example.* In the previous two examples, the maximal violation of a polynomial Bell inequality was attained at a maximally entangled state. Next, consider the expression

$$\begin{aligned}&\lambda(A_2+B_1+B_2-A_1B_1+A_2B_1+A_1B_2+A_2B_2)\\ &-\lambda(A_1)\lambda(B_1)-\lambda(A_2)\lambda(B_1)-\lambda(A_2)\lambda(B_2)-\lambda(A_1)^2-\lambda(B_2)^2\,.\end{aligned}\tag{7.9}$$

Below we show that:

(i) (7.9) is bounded by 3.375 for classical models and spatial quantum models with maximally entangled states, and this bound is obtained by a classical model with a discrete 3-atomic measure;

(ii) (7.9) is bounded by 3.51148 for any quantum commuting model, and this bound is obtained by a spatial quantum model of local dimension 2.

Let

$$\begin{aligned}b=&\varsigma(x_2)+\varsigma(y_1)+\varsigma(y_2)-\varsigma(x_1y_1)+\varsigma(x_2y_1)+\varsigma(x_1y_2)+\varsigma(x_2y_2)\\ &-\varsigma(x_1)\varsigma(y_1)-\varsigma(x_2)\varsigma(y_1)-\varsigma(x_2)\varsigma(y_2)-\varsigma(x_1)^2-\varsigma(y_2)^2\,.\end{aligned}$$

(ii): To solve the optimization problem

$$\sup\ b\ \text{ s.t. }\ x_i^2=1,\, y_j^2=1,\, [x_i,y_j]=0 \text{ for } i,j=1,2,\tag{7.10}$$

we first solve the SDP for the relaxation of (7.10) with $d = 2$ as in (6.6). The output is 3.51148; moreover, the resulting Hankel matrix is flat, and the assumptions of Proposition 6.10 are satisfied. Therefore we can perform the finite-dimensional GNS construction and extract a 4-dimensional quantum commuting model attaining 3.51148. After a unitary basis change, the extracted model is evidently spatial quantum, of local dimension 2, and given by

$$|\psi\rangle = \begin{pmatrix} -\cos\beta\sin\frac{\beta}{3} \\ \cos\beta\cos\frac{\beta}{3} \\ -\sin\beta\cos\frac{2\beta-\alpha}{3} \\ \sin\beta\sin\frac{2\beta-\alpha}{3} \end{pmatrix}, \quad A_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes I, \quad A_2 = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix} \otimes I,$$

$$B_1 = I \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad B_2 = -I \otimes \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}$$

and $\lambda(Y) = \langle\psi| Y |\psi\rangle$ for $\alpha = -4.525$ and $\beta = 2.192$.

(i): If in the optimization problem (7.10) one restricts only to *tracial* states (i.e., $\lambda \in \mathcal{S}(\mathcal{H})$ satisfying $\lambda(uv) = \lambda(vu)$), then its solution gives an upper bound of (7.9) for both classical models and spatial quantum models with a maximally entangled state. Using the relaxation with $d = 2$ in the SDP hierarchy [KMV22, Section 5.3] for the tracial version of (7.10) one obtains an upper bound 3.375. Again, the resulting Hankel matrix is flat, so a maximizing 3-dimensional model with a tracial state can be extracted [KMV22, Section 5.4]. In this model, all the operators commute, so the model is classical, on a probability space of size 3. Once the bound on the size of the probability space is known, we can search for a maximizing classical model exactly, resulting in

$$\rho = \operatorname{diag}\left(\frac{1}{4}, \frac{3}{8}, \frac{3}{8}\right), \quad A_1 = \operatorname{diag}(1, -1, -1), \quad A_2 = \operatorname{diag}(1, 1, -1), \quad B_1 = I, \quad B_2 = A_2$$

and $\lambda(Y) = \operatorname{tr}(\rho Y)$, for which the value of (7.9) is $\frac{27}{8} = 3.375$. Lastly, since ρ has rational entries with denominator 8, the upper bound $\frac{27}{8}$ can also be reached by a spatial quantum model with a maximally entangled state with (possibly non-minimal) local dimension $8 \cdot 3 = 24$.

To complete the picture, let us mention that the maximum of (7.9) for deterministic models is 2.

8. Bell inequalities for network scenarios

As seen in the previous section, a polynomial Bell inequality corresponds to optimizing a state polynomial subject to noncommutative constraints. On the other hand, correlation inequalities for general quantum networks [Fri12, PKRR+19, LGG21, TPKLR22] correspond to optimizing a state polynomial subject to both noncommutative and state constraints.

Following [Fri12], a *correlation* or *network scenario* is given by

(1) a set $[M] = \{1, \ldots, M\}$ of parties;
(2) a set $[S] = \{1, \ldots, S\}$ of sources; and

(3) a relation $\rightsquigarrow$ on $[S] \times [M]$, where $s \rightsquigarrow m$ means that the party m has access to the source s.

Parties can have several inputs (questions) and outputs (answers); for $m \in [M]$ let a_m and b_m be the number of inputs and outputs of m, respectively.

A *(spatial) quantum* model for such a network is given by

(i) (finite-dimensional) Hilbert spaces $\mathcal{H}_{(s,m)}$ for $s \rightsquigarrow m$;

(ii) for each $m \in [M]$, projective-valued measures (PVMs) $(P_{m,i,j})_{j=1}^{b_m}$ for $i = 1, \dots, a_m$ where

$$P_{m,i,j} \in \mathcal{B}\left(\bigotimes_{s:s\rightsquigarrow m} \mathcal{H}_{s,m} \right), \qquad P_{m,i,j} = P_{m,i,j}^* = P_{m,i,j}^2, \quad \sum_{j=1}^{b_m} P_{m,i,j} = I;$$

(iii) density matrices

$$\rho_s \in \mathcal{B}\left(\bigotimes_{m:s\rightsquigarrow m} \mathcal{H}_{s,m} \right)$$

representing states, for $s \in [S]$.

The correlations of this model are

$$p(i_1, j_1, i_2, j_2, \dots, i_M, j_M) = \operatorname{tr}\left(\bigotimes_{s\in[S]} \rho_s \cdot \bigotimes_{m\in[M]} P_{m,i_m,j_m} \right)$$

with a slight abuse of notation, since the tensor factors need to be appropriately ordered.

As in [LGG21, Section II.C] and [RX22, Definition 3.2], a *quantum commuting model* for such a network is given by

(i) a (possibly infinite-dimensional) Hilbert space $\mathcal{H}$;

(ii) for each $m \in [M]$, projective-valued measures (PVMs) $(P_{m,i,j})_{j=1}^{b_m}$ for $i = 1, \dots, a_m$ where

$$P_{m,i,j} \in \mathcal{B}(\mathcal{H}), \qquad P_{m,i,j} = P_{m,i,j}^* = P_{m,i,j}^2, \quad \sum_{j=1}^{b_m} P_{m,i,j} = I,$$

and

$$[P_{m,i,j}, P_{m',i',j'}] = 0 \qquad \text{for } m \neq m';$$

(iii) a state $\lambda \in \mathcal{S}(\mathcal{H})$ satisfying

$$\lambda(Q_1 \cdots Q_\ell) = \lambda(Q_1) \cdots \lambda(Q_\ell) \tag{8.1}$$

whenever each Q_k is in the algebra generated by the PVMs for m_k, and for all $k \neq k'$, $m_k \neq m_{k'}$ and there is no $s \in [S]$ with $m_k \leftsquigarrow s \rightsquigarrow m_{k'}$.

The correlations of this model are

$$p(i_1, j_1, i_2, j_2, \dots, i_M, j_M) = \lambda\left(P_{1,i_1,j_1} P_{2,i_2,j_2} \cdots P_{M,i_M,j_M}\right)$$

Clearly, correlations of spatial quantum models are produced by quantum commuting models. When all operators in a quantum commuting model commute (in which case measurements are given by indicator functions on a probability space, and the state is given by the integration with respect to the probability measure), the model is *classical*.

A (classical/spatial quantum/quantum commuting) polynomial Bell inequality for a network scenario is an upper bound on a polynomial expression in correlations, valid for every (classical/spatial quantum/quantum commuting) model. We can obtain Bell inequalities for quantum commuting models of network scenarios using the SDP hierarchy from Section 6 as follows.

Consider the description of a network scenario as at the beginning of this section, and let $\mathcal{B}$ be a polynomial expression in correlations $p(i_1, j_1, \dots, i_M, j_M)$. For each m, i, j let $x_{m,i,j}$ be a freely noncommuting self-adjoint variable, and let $b \in \mathscr{S}$ be the state polynomial obtained from $\mathcal{B}$ by replacing $p(i_1, j_1, \dots, i_M, j_M)$ with $\varsigma(x_{1,i_1,j_1} \cdots x_{M,i_M,j_M})$.

Corollary 8.1. *Let $\mathcal{B}$ and b be as above. Then $\beta \in \mathbb{R}$ is the smallest constant such that $\mathcal{B} \le \beta$ for every quantum commuting model if and only if β is the output of the state polynomial optimization problem*

$$
\begin{aligned}
&\sup\ b \\
&\text{s.t.}\quad x_{m,i,j}^2 = x_{m,i,j},\ \sum_{j=1}^{b_m} x_{m,i,j} = 1\,, \\
&\qquad x_{m,i,j}x_{m,i,j'} = 0, \quad \text{for } j \neq j',\ [x_{m,i,j}, x_{m',i',j'}] = 0 \quad \text{for } m \neq m'\,, \\
&\qquad \varsigma(w_1 \cdots w_\ell) = \varsigma(w_1) \cdots \varsigma(w_\ell) \quad \text{for } w_k \in \langle x_{m_k,i,j} : i, j \rangle \text{ where } m_k \text{ are distinct} \\
&\qquad\qquad \text{and not sharing sources.}
\end{aligned}
\tag{8.2}
$$

Note that the constraints $x_{m,i,j}x_{m,i,j'} = 0$ for $j \neq j'$ above are redundant, but convenient for reducing the size of SDPs.

In particular, Corollaries 6.1 and 8.1 together yield a convergent SDP hierarchy for optimization of Bell expressions over quantum commuting models of an arbitrary network scenario. By Proposition 6.7, it is easy to see that in the special case of the bilocal scenario (see Section 8.1 below), this SDP hierarchy is equivalent to the one presented in [PKRR+19], and whose convergence was first proved in [RX22]. For a different convergent SDP hierarchy based on the quantum de Finetti theorem, see [LGG21].

8.1. **Bilocal scenario.** In the *bilocal* scenario, there are three parties and two sources, and the middle party shares a source with each of the other parties. In the network notation above, $M = \{1, 2, 3\}$, $S = \{1, 2\}$ and

$$1 \leftsquigarrow 1 \rightsquigarrow 2 \leftsquigarrow 2 \rightsquigarrow 3.$$

Below we provide bounds on quantum commuting violations of some Bell inequalities of classical models for this scenario. We restrict ourselves to 2-output scenarios; note that a two-output PVM $\{P, I - P\}$ can be equivalently given by a binary observable A (namely, $A = 2P - I$). Thus we describe measurements of the first (resp. second; resp. third) party in terms of binary observables A_i (resp. B_i; resp. C_i), as in Section 7.

8.1.1. *Example.* Suppose each of the parties has two inputs and two outputs. In [Cha16], the following inequality for classical models of this bilocal scenario is given:

$$\sqrt{|J_1|}+\sqrt{|J_2|}\leq 2 \tag{8.3}$$

where

$$J_1=\lambda\big((A_1+A_2)B_1(C_1+C_2)\big),\qquad J_2=\lambda\big((A_1-A_2)B_2(C_1-C_2)\big).$$

The inequality (8.3) is equivalent to the four polynomial inequalities

$$-\frac{1}{8}(J_1-\eta_1J_2)^2+\eta_2(\eta_1J_1+J_2)\leq 2 \tag{8.4}$$

for $\eta_i\in\{-1,1\}$; see [Cha16]. Let us bound quantum commuting violations of (8.4) for $\eta_1=\eta_2=1$ (the other cases are similar). Denote

$$j_1=\sum_{i,j\in\{1,2\}}\varsigma(x_iy_1z_j),\quad j_2=\sum_{i,j\in\{1,2\}}(-1)^{i+j}\varsigma(x_iy_2z_j),\quad b=-\frac{1}{8}(j_1-j_2)^2+(j_1+j_2).$$

The maximal bilocal quantum commuting violation of (8.4) is then given by the optimization problem

$$\begin{aligned}\sup\ b\ \text{ s.t. }\ & x_i^2=y_i^2=z_i^2=1,\quad [x_i,y_j]=[y_i,z_j]=[z_i,x_j]=0 \text{ for all } i,j,\\ & \varsigma\big(w_1(x_1,x_2)w_2(z_1,z_2)\big)=\varsigma\big(w_1(x_1,x_2)\big)\varsigma\big(w_2(z_1,z_2)\big) \text{ for all } w_1,w_2.\end{aligned} \tag{8.5}$$

The SDP (6.6) for $d=3$ returns 4, which gives an upper bound for a bilocal quantum commuting violation of (8.4). This bound is attained by a spatial quantum model with

$$\begin{aligned}
&A_1=C_1=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\qquad A_2=C_2=\begin{pmatrix}0&1\\1&0\end{pmatrix},\\
&B_1=\frac{1}{2}\begin{pmatrix}1&1\\1&-1\end{pmatrix}\otimes\begin{pmatrix}1&1\\1&-1\end{pmatrix},\quad B_2=\frac{1}{2}\begin{pmatrix}1&-1\\-1&-1\end{pmatrix}\otimes\begin{pmatrix}1&-1\\-1&-1\end{pmatrix},\\
&\rho_j=|\psi\rangle\langle\psi|\quad\text{for}\quad|\psi\rangle=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}.
\end{aligned}$$

Also, for this bilocal model one has $\sqrt{|J_1|}+\sqrt{|J_2|}=2\sqrt{2}$.

8.1.2. *Example.* In [RBB+16, Appendix F], the authors prove an analog of the I_{3322} inequality for classical bilocal models

$$\sqrt{|J_1|}+\sqrt{|J_2|}\leq\sqrt{|L|} \tag{8.6}$$

where

$$\begin{aligned}
J_1&=\frac{1}{2}\lambda\big((A_1+A_2+A_3+I)B_1(C_1+C_2)\big),\\
J_2&=\frac{1}{2}\lambda\big((A_1+A_2-A_3+I)B_2(C_1-C_2)\big)+\frac{1}{2}\lambda\big((A_1-A_2)B_3(C_1-C_2)\big),\\
L&=4+\lambda(A_1)+\lambda(A_2).
\end{aligned}$$

The inequality (8.6) implies the polynomial inequality

$$2(J_1J_2 + J_1L + J_2L) - J_1^2 - J_2^2 - L^2 \leq 0. \tag{8.7}$$

To find an upper bound for bilocal quantum commuting violations of (8.7), we set up the optimization problem as in the previous example, and the SDP (6.6) for $d = 3$ returns 15.6705. With the current computational limitations, we do not know whether this is the least upper bound. Nonetheless, there certainly exist bilocal quantum violations of (8.7). Concretely, for $\alpha = 1.947$ and $\beta = 1.639$, the spatial quantum model

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad C_1 = A_1, \quad C_2 = A_2,$$

$$B_1 = (\sin\alpha A_2 + \cos\alpha A_3) \otimes \frac{1}{\sqrt{2}}(A_1 + A_2), \quad B_2 = (\sin\alpha A_2 - \cos\alpha A_3) \otimes \frac{1}{\sqrt{2}}(A_1 - A_2),$$

$$B_3 = -A_1 \otimes \frac{1}{\sqrt{2}}(A_1 - A_2),$$

$$\rho_j = |\psi_j\rangle\langle\psi_j| \quad \text{for} \quad |\psi_1\rangle = \frac{\sin\beta}{2}\begin{pmatrix} \sqrt{2} \\ -1 \\ 0 \\ -1 \end{pmatrix} + \frac{\cos\beta}{2}\begin{pmatrix} 0 \\ -1 \\ \sqrt{2} \\ 1 \end{pmatrix}, \quad |\psi_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$$

gives $2(J_1J_2 + J_1L + J_2L) - J_1^2 - J_2^2 - L^2 = 13.3309$.

8.1.3. *Example.* Suppose each of the parties has three inputs and two outputs. In [TPKLR22], an inequality for classical models is given:

$$\frac{1}{3}S - T \leq 3 + 5Z$$

where

$$\begin{aligned} S &= \sum_{i\in\{1,2,3\}} \Big(\lambda(B_iC_i) - \lambda(A_iB_i)\Big), \\ T &= \sum_{\{i,j,k\}=\{1,2,3\}} \lambda(A_iB_jC_k), \\ Z &= \max\Big(\{|\lambda(A_i)|, |\lambda(B_i)|, |\lambda(C_i)| \colon i \in \{1,2,3\}\} \\ &\qquad \cup \{|\lambda(A_iB_j)|, |\lambda(B_iC_j)|, |\lambda(A_iC_j)| \colon i \neq j\} \\ &\qquad \cup \{|\lambda(A_iB_jC_k)| \colon |\{i,j,k\}| \leq 2\}\Big). \end{aligned}$$

In particular, [TPKLR22] focused on the inequality

$$\frac{1}{3}S - T \leq 3 \qquad \text{subject to } Z = 0 \tag{8.8}$$

for classical models, and showed it admits a spatial quantum violation satisfying $Z = 0$ and $\frac{1}{3}S - T = 4$. To provide an upper bound for bilocal quantum commuting violations of (8.8), we set up the optimization problem as before. The SDP (6.6) for $d = 3$ (more precisely, its reduction as in Section 6.4.2 and Section 7.1) contains four PSD blocks with respective size $130, 105, 105, 105$ and 3018 affine constraints. We obtain an upper bound 4.46613 in 1.94s. For $d = 4$, the SDP (6.6) (more precisely, its reduction as in Section 6.4.2 and Section 7.1) contains four PSD blocks with respective size $678, 678, 678, 646$ and 64878 affine constraints. We obtain an upper bound 4.37666 for the quantum commuting violations of (8.8) in 5346s. The corresponding Hankel matrix is not flat and so we cannot certify the optimality of this bound. The SDP (6.6) for $d = 5$ contains four PSD blocks with respective size $3838, 3838, 3838, 3739$ and $1, 352, 093$ affine constraints, which is currently out of reach.

References

[Bar02] Alexander Barvinok. *A course in convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002. 23

[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, 21

[BKP16] Sabine Burgdorf, Igor Klep, and Janez Povh. *Optimization of polynomials in non-commuting variables*. SpringerBriefs in Mathematics. Springer, 2016. 2, 13, 19, 21

[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. 21

[Cha16] Rafael Chaves. Polynomial Bell inequalities. *Phys. Rev. Lett.*, 116(1):010402, 6, 2016. 2, 5, 33

[Cim09] Jakob Cimprič. A representation theorem for archimedean quadratic modules on $*$-rings. *Canad. Math. Bull.*, 52:39–52, 2009. 14, 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

[Fri12] Tobias Fritz. Beyond Bell's theorem: correlation scenarios. *New J. Phys.*, 14(October):103001, 35, 2012. 2, 5, 30

[GdLL19] Sander Gribling, David de Laat, and Monique Laurent. Lower bounds on matrix factorization ranks via noncommutative polynomial optimization. *Found. Comput. Math.*, 19:1013–1070, 2019. 2

[GLS22] Sander Gribling, Monique Laurent, and Andries Steenkamp. Bounding the separable rank via polynomial optimization. *Linear Algebra Appl.*, 648:1–55, 2022. 2

[Hel02] J. William Helton. "Positive" noncommutative polynomials are sums of squares. *Ann. of Math. (2)*, 156(2):675–694, 2002. 2

[HKM12] J. William Helton, Igor Klep, and Scott McCullough. The convex Positivstellensatz in a free algebra. *Adv. Math.*, 231(1):516–534, 2012. 10, 11

[HKMV22] Felix Huber, Igor Klep, Victor Magron, and Jurij Volčič. Dimension-free entanglement detection in multipartite werner states. *Commun. Math. Phys.*, 396(3):1051–1070, 2022. 3

[HM04] J. William Helton and Scott A. McCullough. A Positivstellensatz for non-commutative polynomials. *Trans. Amer. Math. Soc.*, 356(9):3721–3737, 2004. 2, 7, 14

[JNV+21] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. MIP* = RE. *Commun. ACM*, 64:131–138, 2021. 7

[KMP22] Igor Klep, Victor Magron, and Janez Povh. Sparse noncommutative polynomial optimization. *Math. Program.*, 193(2):789–829, 2022. 22, 23, 24

[KMV22] Igor Klep, Victor Magron, and Jurij Volčič. Optimization over trace polynomials. *Ann. Henri Poincaré*, 23(1):67–100, 2022. 3, 7, 13, 16, 21, 30

[KPV21] Igor Klep, James E. Pascoe, and Jurij Volčič. Positive univariate trace polynomials. *J. Algebra*, 579:303–317, 2021. 3, 10

[KS08] Igor Klep and Markus Schweighofer. Connes' embedding conjecture and sums of Hermitian squares. *Adv. Math.*, 217(4):1816–1837, 2008. 7

[KŠ17] Igor Klep and Špela Špenko. Free function theory through matrix invariants. *Canad. J. Math.*, 69(2):408–433, 2017. 7, 8

[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, 10, 12

[KSV22] Igor Klep, Claus Scheiderer, and Jurij Volčič. Globally trace-positive noncommutative polynomials and the unbounded tracial moment problem. *Math. Ann. to appear,* `doi.org/10.1007/s00208-022-02495-5`, 2022. 3, 10

[KVV14] Dmitry S. Kaliuzhnyi-Verbovetskyi and Victor Vinnikov. *Foundations of free noncommutative function theory*, volume 199 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2014. 7, 8

[Las01] Jean-Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2000/01. 2

[LGG21] Laurens T. Ligthart, Mariami Gachechiladze, and David Gross. A convergent inflation hierarchy for quantum causal structures, 2021. preprint `https://arxiv.org/abs/2110.14659`. 5, 30, 31, 32

[LLR08] Jean Bernard Lasserre, Monique Laurent, and Philipp Rostalski. Semidefinite characterization and computation of zero-dimensional real radical ideals. *Found. Comput. Math.*, 8(5):607–647, 2008. 22

[Mar08] Murray Marshall. *Positive polynomials and sums of squares*, volume 146 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2008. 3, 10, 11, 13, 17

[McC01] Scott McCullough. Factorization of operator-valued polynomials in several non-commuting variables. *Linear Algebra Appl.*, 326(1-3):193–203, 2001. 2

[MW23] Victor Magron and Jie Wang. *Sparse polynomial optimization: theory and practice*. Series on Optimization and Its Applications, World Scientific Press, 2023. Accepted for publication. 2, 28

[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. 28

[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. 2, 26

[PHBB17] Victor Pozsgay, Flavien Hirsch, Cyril Branciard, and Nicolas Brunner. Covariance Bell inequalities. *Phys. Rev. A*, 96(6):062128, 13, 2017. 2, 26, 29

[PKRR+19] Alejandro Pozas-Kerstjens, Rafael Rabelo, Łukasz 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. 2, 5, 30, 32

[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, 21

[Pro76] Claudio Procesi. The invariant theory of $n \times n$ matrices. *Adv. Math.*, 19(3):306–381, 1976. 7, 8, 9

[Put93] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.*, 42(3):969–984, 1993. 2, 4, 12

[RBB+16] Denis Rosset, Cyril Branciard, Tomer Jack Barnea, Gilles Pütz, Nicolas Brunner, and Nicolas Gisin. Nonlinear Bell inequalities tailored for quantum networks. *Phys. Rev. Lett.*, 116:010403, Jan 2016. 33

[RX22] Marc-Olivier Renou and Xiangling Xu. Two convergent NPA-like hierarchies for the quantum bilocal scenario, 2022. preprint `https:arXiv.org/abs/2210.09065`. 5, 31, 32

[Sch09] Claus Scheiderer. Positivity and sums of squares: a guide to recent results. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 271–324. Springer, New York, 2009. 3, 13

[SS13] Yurii Savchuk and Konrad Schmüdgen. Unbounded induced representations of $*$-algebras. *Algebr. Represent. Theory*, 16:309–376, 2013. 27, 28

[TPKLR22] Armin Tavakoli, Alejandro Pozas-Kerstjens, Ming-Xing Luo, and Marc-Olivier Renou. Bell nonlocality in networks. *Rep. Progr. Phys.*, 85(5):Paper No. 056001, 41, 2022. 2, 30, 34

[Uff02] Jos Uffink. Quadratic Bell inequalities as tests for multipartite entanglement. *Phys. Rev. Lett.*, 88(23):230406, 4, 2002. 5, 26, 28

[WM21] Jie Wang and Victor Magron. Exploiting term sparsity in noncommutative polynomial optimization. *Comput. Optim. Appl.*, 80(2):483–521, 2021. 22, 25

[WSV12] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of semidefinite programming: theory, algorithms, and applications*, volume 27. Springer Science & Business Media, 2012. 22

Igor Klep: Faculty of Mathematics and Physics, Department of Mathematics, University of Ljubljana, Slovenia

Email address: `igor.klep@fmf.uni-lj.si`

Victor Magron: LAAS-CNRS & Institute of Mathematics from Toulouse, France

Email address: `vmagron@laas.fr`

Jurij Volčič: Department of Mathematics, Drexel University, Pennsylvania

Email address: `jurij.volcic@drexel.edu`

Jie Wang: Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

Email address: `wangjie212@amss.ac.cn`

NOT FOR PUBLICATION

Contents