CERTIFYING OPTIMALITY OF BELL INEQUALITY VIOLATIONS: NONCOMMUTATIVE POLYNOMIAL OPTIMIZATION THROUGH SEMIDEFINITE PROGRAMMING AND LOCAL OPTIMIZATION

TIMOTEJ HRGA, IGOR KLEP, AND JANEZ POVH

Abstract. Bell inequalities are pillars of quantum physics in that their violations imply that certain properties of quantum physics (e.g. entanglement) cannot be represented by any classical picture of physics. In this article Bell inequalities and their violations are considered through the lens of noncommutative polynomial optimization. Optimality of these violations is certified for a large majority of a set of standard Bell inequalities, denoted A2–A89 in the literature. The main techniques used in the paper include: the NPA hierarchy, i.e., the noncommutative version of the Lasserre semidefinite programming (SDP) hierarchies based on the Helton-McCullough Positivstellensatz, the Gelfand-Naimark-Segal (GNS) construction with a novel use of the Artin-Wedderburn theory for rounding and projecting, and nonlinear programming (NLP). A new "Newton chip"-like technique based on conditional expectations for reducing sizes of SDP arising in the constructed polynomial optimization problems is presented. Finally, noncommutative Gröbner bases are exploited to certify when a dual SDP solution does *not* give rise to an optimizer is given.

1. Introduction

Bell inequalities, introduced by Bell in his seminal paper [Bel64], have been instrumental in the quest to experimentally demonstrate the validity of quantum mechanics. Violation of a Bell inequality serves as an indicator for entanglement of a quantum state and implies that a physical interaction cannot be explained via locally causal models from classical physics. Mathematically a Bell inequality is simply a special type of inequality on (eigenvalues of) noncommutative polynomials as sketched in Section 1.1, cf. [PNA10]. Thus finding violations of Bell inequalities requires solving instances of noncommutative polynomial optimization problems.

Polynomial optimization minimizes a polynomial over a semialgebraic set, i.e., a set defined by polynomial inequalities. Since this optimization problem is NP-hard [Lau09], various relaxation schemes are employed. The most successful one is the Lasserre hierarchy [Las01] based on Putinar's Positivstellensatz [Put93], a powerful representation result from real algebraic geometry. Lasserre's hierarchy produces a sequence of semidefinite programming (SDP) relaxations whose optima converge to the optimum of the original problem (under mild natural assumptions). Polynomial optimization and its interplay with real algebraic geometry remain a highly vibrant area, and we refer the reader to e.g. [Las10, HLL09, Sch05, RTAL13, MHL15, WML21, AM19, Nie14, Lau09] and the references therein for further details.

In the *noncommutative* context one considers polynomials in noncommuting variables and their positivity when evaluated at tuples of matrices (or, more generally, operators on Hilbert space). Surprisingly, a noncommutative polynomial is positive (semidefinite) if and only if it is a *sum of hermitian squares* (sos) [Hel02, McC01]. As in the commutative case, one relies on sos to optimize noncommutative polynomials over noncommutative semialgebraic sets, i.e., sets described by noncommutative polynomial inequality constraints. This is done with the noncommutative Lasserre hierarchy [HM04, DLTW08, NPA08, PNA10, CKP12, BKP16], also called the

Date: March 17, 2023.

2020 *Mathematics Subject Classification.* 90C22; 47N10; 13J30.

Key words and phrases. noncommutative polynomial; Bell inequality; violation; Gröbner basis; semidefinite programming; eigenvalue optimization; GNS construction; Artin-Wedderburn theory.

IK was supported by the Slovenian Research Agency program P1-0222 and grants J1-2453, J1-3004, and N1-0057.

TH was supported by the young researcher programme MR+ of the Slovenian Research Agency.

JP was supported bt the Slovenian Research Agency program P2-0162 and grants J1-1691, J1-2453, J2-2512, J5-2552, and N1-0071.

Navascués-Pironio-Acín (NPA) hierarchy. The impetus for progress has been provided by linear systems theory at first ([DOHMP09, Hel02]), and lately quantum information theory has been a major driver of progress [DLTW08, NPA08, BFS16, GdLL19]. `NCSOStools` [CKP11, BKP16] can compute lower bounds on minimal eigenvalues or traces of noncommutative polynomial objective functions over noncommutative semialgebraic sets; also see the Julia library `TSSOS` of Magron and Wang [MW23, Appendix B].

We next survey how Bell inequalities give rise to noncommutative optimization problems, and our contribution including a reader's guide is then given in Section 1.2.

1.1. **From Bell inequalities to noncommutative polynomial optimization.** Suppose we have a system consisting of two subsystems, e.g. two players Alice and Bob. It operates as follows: Alice and Bob receive questions from the finite sets of questions S and T, respectively, and provide answers from the finite sets of answers A and B, respectively. We consider conditional joint probabilities

$$P(a,b|s,t) := P(\text{Alice answers } a,\ \text{Bob answers } b|\text{Alice is asked } s,\ \text{Bob is asked } t),$$

called *correlations* in the quantum literature [Gis09].

If Alice and Bob use deterministic strategies (so-called classical configuration of the system [AIIS05]), which means that

$$\begin{aligned} P_A(a|s) &:= P(\text{Alice answers } a|\text{Alice is asked } s) \in \{0,1\}\\ P_B(b|t) &:= P(\text{Bob answers } b|\text{Bob is asked } t) \in \{0,1\}, \end{aligned}$$

for all $s \in S, t \in T, a \in A, b \in B$, then we have deterministic correlations [GdLL18] and $P(a,b|s,t) = P_A(a|s)P_B(b|t) \in \{0,1\}$ for all a,b,s,t. If they use independent probabilistic strategies (i.e., for each $s \in S$ Alice decides which answer from A to provide based on some probability distribution p_s, i.e., $P_A(a|s) = p_s(a)$ and the same for Bob, and these distributions are independent of each other), we have $P(a,b|s,t) = p_s(a)p_t(b) \in [0,1]$. If their strategy is based on some shared randomness, i.e., if the pair of answers (a,b) to the pair of questions (s,t) is determined randomly according to some probability distribution $p_{s,t}$, the resulting correlations are called *classical correlations* and are equal to $P(a,b|s,t) = p_{s,t}(a,b)$. That is, classical correlations are convex combinations of deterministic correlations and they represent the so-called *local hidden variable model* [Gis09, AIIS05].

Bell inequalities are linear inequalities in the correlations $P(a,b|s,t)$ and in the marginal probabilities $P_A(a|s)$ and $P_B(b|t)$, that are valid for all classical correlations and define the so-called Bell polytope [Gis09, AIIS05]. The tight Bell inequalities are the facet-defining inequalities for this polytope. However, quantum systems can violate some of them, as has been confirmed by several experiments, see e.g. [RKM+01]. One of the most famous tight Bell inequalities violated by quantum systems is the Clauser-Horne-Shimony-Holt (CHSH) inequality [CHSH69], which classically reads for the case when we have $S = T = A = B = \{0,1\}$ as follows [PV09]:

$$-P_A(1|0) - P_B(1|0) + P(1,1|0,0) + P(1,1|0,1) + P(1,1|1,0) - P(1,1|1,1) \le 0. \tag{1.1}$$

A (quantum) system violates this inequality if the left-hand side is positive, which is equivalent to the condition that the negative value of the left-hand side is negative. To be compliant with the rest of the paper, we introduce the *violation* of the Bell inequality (1.1) as the negative value of the left-hand side of (1.1) if this value is negative, i.e., violation $:= P_A(1|0) + P_B(1|0) - P(1,1|0,0) - P(1,1|0,1) - P(1,1|1,0) + P(1,1|1,1)$, if this value is negative. Tsirelson [Tsi80] proved the largest (most negative) violation for the CHSH inequality is $-(\sqrt{2}-1)/2$.

Quantum systems differ from classical systems in that Alice and Bob follow a quantum strategy, which means that they share a bipartite quantum state ψ and they answer the questions s,t by performing quantum measurements on their part of the quantum state. A standard representation of quantum measurements is the commuting model, where these measurements are represented by projector-valued measures, or PVM for short, which are defined to be collections of bounded operators on a separable Hilbert space $\cup_{s\in S}\{X_s^a \mid a \in A\}$ and $\cup_{t\in T}\{Y_t^b \mid b \in B\}$, satisfying [NPA08]:

$$X_s^aY_t^b = Y_t^bX_s^a, \qquad X_s^aX_s^a = X_s^a, \qquad Y_t^bY_t^b = Y_t^b, \qquad X_s^aX_s^{a'} = 0, \qquad Y_t^bY_t^{b'} = 0,$$

$$\sum_a X_s^a = I, \qquad \sum_b Y_t^b = I$$

for all $a, b, s, t, a' \neq a, b' \neq b$.

Based on PVM and a shared quantum state ψ, Alice and Bob return the pair of answers (a, b) to the pair of questions (s, t) with probability

$$P(a, b|s, t) = \psi^T X_s^a Y_t^b \psi.$$

Likewise, $P_A(a|s) = \psi^T X_s^a \psi$ and $P_B(b|t) = \psi^T Y_t^b \psi$.

To look for the largest (most negative) violation of some tight Bell inequality

$$\sum_{a,s} c_{a,s} P_A(a|s) + \sum_{b,t} c_{b,t} P_B(b|t) + \sum_{a,b,s,t} c_{a,b,s,t} P(a, b|s, t) \geq d,$$

we have to minimize the left hand side.

For the CHSH example, we have to minimize

$$\psi^\star(X_0^1 + Y_0^1 - X_0^1 Y_0^1 - X_0^1 Y_1^1 - X_1^1 Y_0^1 + X_1^1 Y_1^1)\psi$$

subject to $X_i^1 X_i^1 = X_i^1$, and $Y_i^1 Y_i^1 = Y_i^1$, for $i = 0, 1$, and $X_i^1 Y_j^1 = Y_j^1 X_i^1$, for all $i, j = 0, 1$, where ψ is a complex unit vector. Note that we eliminated X_i^0 and Y_i^0, $i = 0, 1$ since they can be expressed as $X_i^0 = I - X_i^1$ and $Y_i^0 = 1 - Y_i^1$ and they in this form satisfy the constraints from above. In purely mathematical terms, to establish the largest violation we are looking for a minimum eigenvalue of the non-commutative polynomial $X_0^1 + Y_0^1 - X_0^1 Y_0^1 - X_0^1 Y_1^1 - X_1^1 Y_0^1 + X_1^1 Y_1^1$ subject to projection constraints and commutativity of X_i^1 with Y_j^1, which is an instance of the optimization problems studied in noncommutative polynomial optimization [BKP16].

1.2. **Contribution and reader's guide.** In this article we apply recent progress and develop new tools for noncommutative polynomial optimization to certify optimality of violations for a large set of standard Bell inequalities obtained from the quantum physics literature [IIA06, AIIS05, PV09, PV10]. Following [IIA06, PV09] the 88 problems are denoted A2–A89. These are optimization problems where the objective function is a quadratic noncommutative polynomial in between 4 and 10 variables.

In the following, G announces a paragraph that is mostly a guide to the paper, and C denotes a paragraph describing a new contribution. The manuscript is organized as follows.

(G) Section 2 introduces the basic concepts, and Section 3 is mostly expository. Its main purpose is to fix notation, terminology and to keep the presentation self-contained. We specialize the NPA hierarchy of SDP relaxations [DLTW08, NPA08, PNA10, BKP16], i.e., the implementation of the Helton-McCullough noncommutative Positivstellensatz [HM04] to the case of noncommutative polynomial optimization problems arising from Bell inequalities. For this we formally define the Bell algebra in Section 3.1 as the appropriate algebraic object in which to study the violation of Bell inequalities; Section 3.3 explains how we efficiently implement computations with the Bell algebra in Matlab based on the results discussed in Section 3.2. This enables us to solve the third step of the NPA hierarchy for the standard Bell inequalities A2–A89 already using standard SDP solvers.

(C) To further reduce the size of the SDPs constructed in the Bell algebra we employ so-called *SOS conditional expectations* [SS13] in Section 3.5. This reduction is lossless and is the appropriate analog of the Newton polytopes in commutative polynomial optimization [Rez78, BPT13] and Newton chips in the freely noncommutative setting [BKP16].

(C) Our next contribution is in Section 3.4, where we employ a regularization method with which we can for the first time reliably solve the fourth step of the NPA hierarchy for all of the standard Bell inequalities, and even the fifth step for several of them. The regularization method is well-known but has, to the best of our knowledge, not yet been successfully applied to (noncommutative) polynomial optimization problems.

(G)&(C) Section 4 then turns to certifying optimality of Bell inequality violations. For this we apply the flatness rank condition from operator theory championed first in the commutative by Curto-Fialkow [CF00] and in the noncommutative by McCullough [McC01]; cf. [PNA10]. In Section 4.1 we recall how this flat rank condition on the solution of the dual SDP in the NPA hierarchy can be exploited to extract optimizers via the so–called Gelfand-Naimark-Segal (GNS), and thus certify optimality of a Bell inequality violation. Our contribution here is merely to observe that contrary to widespread belief flatness is in fact a useful tool for noncommutative polynomial optimization. Indeed, for more than 70% of the A2–A89 problems the hierarchy becomes flat quickly (say in the second, third or fourth step of the SDP hierarchy), thus leading to a proof of optimality of the violation.

(C) Flatness typically brings with it numerical challenges. Section 4.2 exploits the concept of semisimplicity from noncommutative algebra [Lam13] and uses the Artin-Wedderburn theory [Lam13, Chapter 1] for rounding and projecting in order to improve the accuracy of solutions obtained from the GNS construction in Section 4.1.

(C) For our final contribution, in an opposite direction, we explain how noncommutative Gröbner bases can be used to certify when the solution to a dual SDP from the hierarchy does *not* give rise to an optimizer; see Section 4.3. This is a noncommutative variant of the technique called recursive generation (RG) in classical moment problems [CF98].

(G) In Section 5 we employ standard nonlinear programming (NLP) to produce upper bounds on Bell inequality violations. The NPA hierarchy produces lower bounds on a Bell inequality violation; often flatness can be used to extract optimizers and thus prove sharpness. In many other cases the upper bounds produced with NLP coincide with those from NPA, thus again certifying optimality. Together with flatness this enables us to certify optimality of 86% of the A2–A89 Bell inequalities.

(G) Numerical results with a detailed description of data are presented in Section 6.

2. Preliminaries

In this section we present basic notation and give some preliminaries.

2.1. **Basic algebraic notation.** Let $\underline{X} = (X_1, \dots, X_m)$ and $\underline{Y} = (Y_1, \dots, Y_n)$ be two tuples of noncommuting variables and denote by $\langle \underline{X} \rangle, \langle \underline{Y} \rangle$ the corresponding sets (monoids) of words generated by X and Y, respectively. Then $\mathbb{R}\langle \underline{X} \rangle$ and $\mathbb{R}\langle \underline{Y} \rangle$ are the corresponding free algebras of noncommutative (nc) polynomials. To study Bell inequalities arising from bipartite states in quantum physics we form the **bipartite free algebra** $\mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle = \mathbb{R}\langle \underline{X} \rangle \otimes \mathbb{R}\langle \underline{Y} \rangle$. Elements of $\mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle$ are of the form

$$f = \sum_{u \in \langle \underline{X} \rangle} \sum_{w \in \langle \underline{Y} \rangle} a_{u,w} uw, \quad a_{u,w} \in \mathbb{R}, \tag{2.1}$$

where the finite sums are over all words in $\langle \underline{X} \rangle$ and $\langle \underline{Y} \rangle$, respectively. We call an element of the form $a_{u,w}uw$ with $a_{u,w} \neq 0$ a **monomial**. Its **degree** is $|u|+|w|$, and the degree of f in (2.1) is the largest degree of a monomial appearing in f. We denote by $\mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle_d$ the set of all elements of degree $\leq d$. Note that by the definition of the bipartite algebra, the words from $\langle \underline{X} \rangle$ and $\langle \underline{Y} \rangle$ commute. The algebra $\mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle$ also comes equipped with the **involution** $\star$, which fixes $\mathbb{R} \cup \{\underline{X}, \underline{Y}\}$ element-wise and reverses words from $\langle \underline{X} \rangle, \langle \underline{Y} \rangle$. Therefore, $f^\star = \sum_{u \in \langle \underline{X} \rangle} \sum_{w \in \langle \underline{Y} \rangle} a_{u,w} u^\star w^\star$, for f as in (2.1), and $\operatorname{Sym} \mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle = \{ f \in \mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle \mid f = f^\star \}$ denotes the set of all **symmetric elements**.

Let $\mathbf{V}_d$ be the (column) vector of all words $uw \in \mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle$ with $u \in \langle \underline{X} \rangle$, $w \in \langle \underline{Y} \rangle$ of degree $\leq d$ sorted w.r.t. the graded lexicographic order. With a slight abuse of notation, we will write $u \in \mathbf{V}_d$, if u is an entry of $\mathbf{V}_d$.

Every $f \in \mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle$ of degree $\leq 2d$ can be written (non-uniquely) as $f = \mathbf{V}_d^\star G \mathbf{V}_d$, where G is a real matrix, called a **Gram matrix** of f. If $f = f^\star$ then G can (and generally will) be chosen to be symmetric. Note that the length of $\mathbf{V}_d$ is $N := N(m,n,d) = \dim \mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle_d$,

$$N = \sum_{\ell=0}^{d} \sigma(m,\ell) n^{d-\ell} = \begin{cases} \frac{(n-1)m^{d+2} - mn^{d+2} + n\left(n^{d+1}-1\right)+m}{(m-1)(n-1)(m-n)} & m \neq n \\ \frac{-\left((d+2)m^{d+1}\right)+(d+1)m^{d+2}+1}{(m-1)^2} & m = n \end{cases},$$

where m is the number of X variables, n is the number of Y variables and $\sigma(m,\ell) = \frac{m^{\ell+1}-1}{m-1}$ is the number of different words (monomials) of length $\leq \ell$ in $m > 1$ non-commuting variables; if $m = 1$, we have $\sigma(1,\ell) = \ell + 1$.

An **ideal** $\mathcal{I}$ of an algebra $\mathcal{A}$ is a vector subspace such that $\mathcal{A}\cdot\mathcal{I}\cdot\mathcal{A}\subseteq\mathcal{I}$. In particular, given a subset $S\subset\mathcal{A}$ the ideal generated by S is

$$\mathcal{I}_S=\Big\{\sum_{i,j}^{\text{finite}} a_{ij}s_ib_{ij}\mid a_{ij},b_{ij}\in\mathcal{A},\ s_i\in S\Big\}.$$

Observe that $\mathbb{R}\langle \underline{X}\leftrightarrow\underline{Y}\rangle$ is the quotient ring $\mathbb{R}\langle\underline{X},\underline{Y}\rangle/\mathcal{I}_S$, where S is the set of all commutators $[X_i,Y_j]$, $i=1,\dots,m$, $j=1,\dots,n$.

By $\mathbb{S}_d$ we denote the vector space of real symmetric $d\times d$ matrices and by $\mathbb{S}_d^{\succeq 0}$ the cone of (real) positive semidefinite $d\times d$ matrices. Occasionally we will need to use complex matrices as well; then we use $\mathbb{S}_d(\mathbb{C})$ and $\mathbb{S}_d(\mathbb{C})^{\succeq 0}$ for the sets of complex hermitian $d\times d$ matrices and its subset of positive semidefinite matrices, respectively.

2.2. **Positivity and sum of squares in $\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle$.** As a stepping stone towards our study of Bell inequality violations, we consider sums of squares and positivity for noncommutative polynomials from the bipartite free algebra $\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle$.

2.2.1. *Sum of squares (sos) in $\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle$.* An element of the form $f^\star f$ is called a (hermitian) **square**, and our main interest is in **sums of squares** (**sos**). An element $f\in\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle$ is sos, if $f\equiv\sum_i g_i^\star g_i$ for some $g_i\in\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle$. With the help of Gram matrices one can detect membership in the sos cone using semidefinite programming (SDP):

Proposition 2.1. *An element $f\in\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle_{2d}$ is sos iff there is $G\succeq 0$ such that*

$$f\equiv\mathbf{V}_d^\star G\mathbf{V}_d. \tag{2.2}$$

Proof. If $f\equiv\sum_i g_i^\star g_i$ is sos, then $\deg g_i\le d$ for all i as the highest degree terms cannot cancel. Hence we can write $g_i=G_i^T\mathbf{V}_d$, where G_i^T is the (row) vector consisting of the coefficients of g_i. Then $g_i^\star g_i=\mathbf{V}_d^\star G_iG_i^T\mathbf{V}_d$ and by setting $G:=\sum_i G_iG_i^T$, (2.2) holds.

Conversely, given a positive semidefinite $G\in\mathbb{R}^{N\times N}$ of rank r satisfying (2.2), write $G=\sum_{i=1}^r G_iG_i^T$ for $G_i\in\mathbb{R}^{N\times 1}$, where $N=N(m,n,d)$ is the length of $\mathbf{V}_d$, defined above. By defining $g_i:=G_i^T\mathbf{V}_d$, we get $f\equiv\sum g_i^\star g_i$. ☐

Therefore, for every $f\in\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle_{2d}$ we can check whether f is sos by solving an instance of semidefinite programming problem, which can be formulated as follows:

$$\begin{aligned}\inf\quad & \langle C,G\rangle\\ \text{s.t.}\quad & f_w=\sum_{u,v\in\mathbf{V}_d,\ u^\star v\equiv w} G_{u,v},\ \forall \text{ entries } w \text{ of } \mathbf{V}_{2d}\\ & G\succeq 0,\end{aligned} \tag{2.3}$$

The matrix C in the objective function could be zero matrix, in which case we tend to get the highest rank solution G. If we want a solution G with a small rank (to obtain a shorter sos decomposition), we can choose for C the identity matrix, i.e., we minimize the trace, which is a a commonly used heuristic for matrix rank minimization [RFP10].

2.2.2. *Evaluation and positivity of elements in $\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle$.* By an **evaluation** of $f\in\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle$ we mean its image under a **$\star$-representation** of $\mathbb{R}\langle\underline{X}\leftrightarrow\underline{Y}\rangle$. Concretely, this means taking a Hilbert space $\mathcal{H}$ and self-adjoint bounded operators x_i,y_j on $\mathcal{H}$ such that

$$[x_i,y_j]=0\quad\text{for all } i,j, \tag{2.4}$$

and then evaluating $f(\underline{x},\underline{y})$. We say f is **positive** if all its evaluations are positive semidefinite.

An important thought not always sufficiently rich class of representations is obtained by considering finite dimensional Hilbert spaces $\mathcal{H}$ in which case $\underline{x},\underline{y}$ are tuples of hermitian matrices satisfying (2.4).

Proposition 2.2. *Suppose $x_i,y_j\in\mathbb{S}_d(\mathbb{C})$ satisfy (2.4). Then for some δ,ϵ we have $d=\delta\varepsilon$, and up to unitary equivalence,*

$$x_i=I_\delta\otimes\xi_i,\quad y_j=\gamma_j\otimes I_\varepsilon \tag{2.5}$$

for some $\xi_i \in \mathbb{S}_\varepsilon(\mathbb{C}), \gamma_j \in \mathbb{S}_\delta(\mathbb{C})$.

Proof. This is well-known; see [DLTW08, Appendix A] for an elementary proof. □

Unlike in the freely noncommutative setting ([Hel02, McC01]), positivity in $\mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y}\rangle$ is *not* equivalent to being sos. This fails already with $n = m = 1$ in which case $\mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y}\rangle = \mathbb{R}[X_1, Y_1]$ is just the algebra of bivariate commutative polynomials and positivity is simply positivity on $\mathbb{R}^2$.

3. Detecting positivity in the Bell algebra via a SDP hierarchy

We now turn our attention to the violation of Bell inequalities. In this section we explain the standard specialization of the noncommutative Lasserre hierarchy (also called the NPA hierarchy after [NPA08]) for the case of Bell inequalities. The hierarchy is based on Proposition 3.1 below (see Corollary 3.3) and gives rise to a sequence of SDPs converging to the true minimum.

It is convenient to consider Bell inequalities in the Bell algebra, formally defined in Section 3.1. Section 3.2 presents the theoretical underpinning of the SDP hierarchy, and Section 3.3 discusses practical implementation of the resulting SDPs. Since the size of the SDPs can grow large quickly, we present in Section 3.4 a regularization method to solve larger SDPs. This enables us to compute, for the first time, fourth levels of the hierarchy for all of the Bell inequalities A2–A82.

3.1. Bell algebra.

We consider the bipartite **Bell algebra** $\mathbb{B}_2 = \mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y}\rangle/\mathcal{I}_B$, where

$$B = \{X_i^2 - X_i, Y_j^2 - Y_j \mid i = 1, \dots, m,\ j = 1, \dots, n\}.$$

Equivalently,

$$\mathbb{B}_2 = \mathbb{R}\langle \underline{X}, \underline{Y}\rangle / \big([X_i, Y_j], X_i^2 - X_i, Y_j^2 - Y_j \mid i = 1, \dots, m,\ j = 1, \dots, n\big).$$

Elements of $\mathbb{B}_2$ are of the form (2.1), but the sums are now over words u, w in $\langle \underline{X}\rangle, \langle \underline{Y}\rangle$, respectively, *without repetitions*.

The involution and degree function on $\mathbb{B}_2$ are inherited from the one on $\mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y}\rangle$, so we can talk about degrees, symmetric elements and (sums of hermitian) squares.

We can evaluate elements f of $\mathbb{B}_2$ in Hilbert spaces $\mathcal{H}$.([1]) Here only tuples of self-adjoint **projections** x_i, y_j on $\mathcal{H}$ are admissible. That is, the self-adjoint bounded operators x_i, y_j on $\mathcal{H}$ must satisfy

$$[x_i, y_j] = 0, \quad x_i^2 = x_i, \quad y_j^2 = y_j \quad \text{for all } i, j. \tag{3.1}$$

We call $f \in \mathbb{B}_2$ **positive** if all its evaluations $f(\underline{x}, \underline{y})$ are positive semidefinite.

3.2. Positivity and sum of squares in the Bell algebra.

We let $\mathbb{V}_d$ denote the vector of all products uw, where u, w are words in $\langle \underline{X}\rangle, \langle \underline{Y}\rangle$, respectively, without repetitions ordered w.r.t. the graded lexicographic order. The entries of $\mathbb{V}_d$ form a basis of $(\mathbb{B}_2)_d$. The number of entries from $\mathbb{V}_d$, having only X_i letters ($i = 1, \dots, m$) is $\hat{\sigma}(m, d) := 1 + \sum_{i=1}^{d} m(m-1)^{i-1}$. Basic calculus shows that $\hat{\sigma}(m, d) = \frac{m(m-1)^d - 2}{m-2}$ if $m \neq 2$. Otherwise, we have $\hat{\sigma}(2, d) = 1 + 2d$. The commutativity between X_i and Y_j implies that the length of $\mathbb{V}_d$ is equal to $\tau(m, n, d) :=$

[1] Our setup is real. That is, we work with polynomials with real coefficients and mostly with real Hilbert spaces. The real framework is more convenient in optimization since the standard SDP solvers normally only accept real-valued problems. While this is contrary to the usual physics setup, where one allows complex Hilbert spaces, it really comes with no loss of generality. Namely, every complex Hilbert space isometrically embeds into a real Hilbert space thus allowing us to model "complex" problems with real Hilbert spaces.

$\dim(\mathbb{B}_2)_d = \hat\sigma(m,d) + \sum_{\ell=1}^{d} \hat\sigma(m,d-\ell)n(n-1)^{\ell-1}$,

$$\tau(m,n,d) = \begin{cases} \frac{m^2(n-2)(m-1)^d + m\left(4-n^2(n-1)^d\right) + 2n\left(n(n-1)^d-2\right)}{(m-2)(n-2)(m-n)} & m \neq 2 \neq n,\ m \neq n, \\ \frac{-n^2(n-1)^d + (n-2)n((d+2)n-2)(n-1)^{d-1}+4}{(n-2)^2} & m = n \neq 2, \\ \frac{n^2(n-1)^d - 4d(n-2) - 4n + 4}{(n-2)^2} & m = 2 \neq n, \\ \frac{m^2(m-1)^d - 4d(m-2) - 4m + 4}{(m-2)^2} & m \neq 2 = n, \\ 1 + 2d + 2d^2 & m = n = 2. \end{cases}$$

As before, every $f \in (\mathbb{B}_2)_{2d}$ has a **Gram matrix** G satisfying $f \equiv \mathbb{V}_d^\star G \mathbb{V}_d$. An element $f \in \mathbb{B}_2$ is a sum of degree d squares iff there is a positive semidefinite $G \succeq 0$ with $f \equiv \mathbb{V}_d^\star G \mathbb{V}_d$ (cf. Proposition 2.1). The set of all such sums of degree d squares is denoted by Σ^2_{2d}, and $\Sigma^2 = \cup_{d\in\mathbb{N}} \Sigma^2_{2d}$ is the set of all sums of squares in $\mathbb{B}_2$. However, in sharp contrast with the bifree setting in $\mathbb{R}\langle \underline{X} \leftrightarrow \underline{Y} \rangle$, sums of squares decompositions in $\mathbb{B}_2$ do not come with degree bounds. This follows from recently established quantum complexity results and the refutation of Connes' embedding conjecture [JNV+21].

To handle quantum Bell inequalities we need to be able to certify positivity of elements of $\mathbb{B}_2$. This is done with a variant of the NPA hierarchy [NPA08] whose theoretical underpinning is the archimedean Positivstellensatz of Helton & McCullough:

Proposition 3.1 ([HM04])**.** *For $f \in \mathbb{B}_2$ the following are equivalent:*

(i) *$f(\underline{x},\underline{y}) \succeq 0$ for all evaluations in Hilbert spaces $\mathcal{H}$;*
(ii) *for all $\varepsilon > 0$, $f + \varepsilon$ is sos in $\mathbb{B}_2$.*

It is important to note that in addition to not having degree bounds in (ii), the ε is needed in general, too.

Proof. To apply [HM04], we need to show that the cone of sums of squares Σ^2 in $\mathbb{B}_2$ is archimedean. Since the set of Σ^2-bounded elements,

$$H = \{f \in \mathbb{B}_2 \mid \exists \eta \in \mathbb{N} : \eta - f^\star f \in \Sigma^2\}$$

is a $*$-subalgebra [Vid59] of $\mathbb{B}_2$, it suffices to show $X_j \in H$. But this is clear since

$$1 - X_j^2 = 1 - X_j = (1 - X_j)^2 \in \Sigma^2. \qquad \square$$

Remark 3.2. An important thought not always sufficiently rich class of representations is obtained by considering finite dimensional Hilbert spaces $\mathcal{H}$ in which case $\underline{x},\underline{y}$ are real symmetric matrices (of the form described in Proposition 2.2 with the additional property that x_i, y_j are projections). However, due to the recently proved failure of the Connes' embedding conjecture [JNV+21, Fri12, JNP+11], there are $f \in \mathbb{B}_2$ whose evaluations on tuples of symmetric projection matrices are positive but they are not positive on infinite dimensional Hilbert spaces. That is, item (i) in Proposition 3.1 is not equivalent to

(i)' $f(\underline{x},\underline{y}) \succeq 0$ for all evaluations in finite-dimensional Hilbert spaces $\mathcal{H}$.

It is a major open problem to find certificates for (i)'. We shall sidestep this issue and focus on evaluations in arbitrary Hilbert spaces, where Proposition 3.1 applies.

Proposition 3.1 yields the following specialization of the NPA hierarchy.

Corollary 3.3. *Let $f \in \mathbb{B}_2$. Consider the optimization problem*

$$\lambda_{\min}(f) := \inf\{\lambda_{\min} f(\underline{x},\underline{y}) \mid \underline{x},\underline{y} \text{ tuples of self-adjoint operators satisfying } (3.1)\}, \tag{3.2}$$

where $\lambda_{\min} f(\underline{x},\underline{y})$ is the smallest eigenvalue of the evaluation $f(\underline{x},\underline{y})$.[2] *For $d \in \mathbb{N}$ we form the d-th order relaxation for $\lambda_{\min}(f)$:*

$$\lambda_d := \sup\{\lambda \mid f - \lambda \in \Sigma^2_{2d}\}. \tag{3.3}$$

[2]More precisely, this is the smallest number λ in the (compact and real) spectrum of $f(\underline{x},\underline{y})$, i.e., for any real number $\mu < \lambda$, $f(\underline{x},\underline{y}) - \mu$ is invertible.

Then λ_d increase and converge to $\lambda_{\min}(f)$ as $d \to \infty$.

The problem (3.3) is a semidefinite program which can be stated as

$$\begin{array}{ll} \sup & f_1 - \langle E_{11}, G\rangle \\ \text{s.t.} & f - f_1 \equiv \mathbb{V}_d^\star \left(G - \langle E_{11}, G\rangle E_{11}\right) \mathbb{V}_d \\ & G \succeq 0, \end{array} \tag{3.4}$$

where f_1 denotes the constant term of f and E_{11} is the matrix unit with a one in the $(1,1)$ entry.

The dual semidefinite program corresponding to (3.4) can be derived in different ways. We follow the procedure in [BKP16, Section 4.2]. Note that Σ_{2d}^2 is a convex cone and its dual can be represented as

$$\begin{aligned} (\Sigma_{2d}^2)^\vee &:= \{L : \mathbb{B}_2 \to \mathbb{R} \mid L \text{ linear and } L(g^\star g) \geq 0 \ \forall g \in (\mathbb{B}_2)_d\} \\ &= \{H \in \mathbb{S}_{\tau(m,n,d)} \mid H \succeq 0,\ H_{1,1} = 1,\ H_{u,v} \ = \ H_{p,q}, \text{ for all } u,v,p,q \in \mathbb{V}_d : u^\star v \equiv p^\star q\}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \lambda_d = \sup_{f-\lambda \in \Sigma_{2d}^2} \lambda \ &= \ \sup_{\lambda} \inf_{L \in (\Sigma_{2d}^2)^\vee} (\lambda + L(f-\lambda)) \\ &\leq \inf_{L \in (\Sigma_{2d}^2)^\vee} \sup_{\lambda} (\lambda + L(f-\lambda)) = \inf_{L \in (\Sigma_{2d}^2)^\vee} \big(L(f) + \sup_{\lambda}(\lambda - L(\lambda))\big) \\ &= \inf_{L \in (\Sigma_{2d}^2)^\vee} \big(L(f) + \sup_{\lambda} \lambda(1 - L(1))\big) \\ &= \inf\{L(f) \mid L \in (\Sigma_{2d}^2)^\vee,\ L(1) = 1\}. \end{aligned}$$

The inequality in the second line is due to the exchange of sup and inf. The optimization problem in the last line is an SDP, which can be reformulated as

$$\begin{array}{lll} \inf & \langle G_f, H\rangle & \\ \text{s.t.} & H_{1,1} \ = \ 1 & \\ & H_{u,v} \ = \ H_{p,q}, & \text{for all } u,v,p,q \in \mathbb{V}_d : u^\star v \equiv p^\star q \\ & H \succeq 0, & \end{array} \tag{3.5}$$

where the matrix G_f is a symmetric Gram matrix for f.

Proposition 3.4. *There is a linear functional $L : (\mathbb{B}_2)_{2d} \to \mathbb{R}$ that is strictly positive in the sense that $L(p^\star p) > 0$ for all $0 \not\equiv p \in (\mathbb{B}_2)_d$.*

Proof. For this proof it is beneficial to present an alternative viewpoint to the Bell algebra $\mathbb{B}_2$, namely as a group algebra. Consider the following linear change of variables:

$$\Delta_i := 2X_i - 1, \quad \Gamma_j := 2Y_j - 1 \tag{3.6}$$

Then

$$\Delta_i^\star = \Delta_i, \quad \text{and} \quad \Delta_i^2 = 1 \tag{3.7}$$

and likewise for Γ_j. Thus the $*$-subalgebra $\mathcal{A}_X$ of $\mathbb{B}_2$ generated by X_i is the same as that generated by Δ_i. Since the Δ_i do not satisfy any other relations besides (3.7), this subalgebra is isomorphic to the group algebra $\mathbb{R}[\mathfrak{G}_m]$, where $\mathfrak{G}_m$ denotes the free product of m (= number of X variables) copies of $\mathbb{Z}/2\mathbb{Z}$. Similarly, the $*$-subalgebra $\mathcal{A}_Y$ of $\mathbb{B}_2$ generated by Y_j is isomorphic to the group algebra $\mathbb{R}[\mathfrak{G}_n]$. Since $\mathcal{A}_X, \mathcal{A}_Y$ together generate $\mathbb{B}_2$ and they commute, we have

$$\mathbb{B}_2 = \mathcal{A}_X \otimes \mathcal{A}_Y \cong \mathbb{R}[\mathfrak{G}_m] \otimes \mathbb{R}[\mathfrak{G}_n] \cong \mathbb{R}[\mathfrak{G}_m \times \mathfrak{G}_n].$$

For simplicity, let $\mathfrak{G} := \mathfrak{G}_m \times \mathfrak{G}_n$.

We now claim that $L : \mathbb{R}[\mathfrak{G}] \to \mathbb{R}$,

$$\sum_{g \in \mathfrak{G}} a_g g \mapsto a_e,$$

where $e \in \mathfrak{G}$ denotes the identity element, is a strictly positive linear functional. It is obviously linear. Given $f = \sum_{g\in\mathfrak{G}} a_g g \in \mathbb{R}[\mathfrak{G}]$, we have $f^\star = \sum_{g\in\mathfrak{G}} a_g g^{-1}$, and thus

$$f^\star f = \sum_{g,h\in\mathfrak{G}} a_g a_h g^{-1}h.$$

The constant term of $f^\star f$, i.e., $L(f^\star f)$, thus equals

$$\sum_{g\in\mathfrak{G}} a_g^2$$

and is nonzero whenever $f \neq 0$. $\square$

This proposition immediately implies a zero duality gap for (3.4)–(3.5), since the constructed strictly positive L yields an interior point for the feasibility set of (3.5).

Corollary 3.5. *The primal-dual pair of SDPs* (3.4)–(3.5) *has zero duality gap.*

In the next section we explain how this SDP is efficiently constructed and solved.

3.3. **Construction of the SDP hierarchy.** In this subsection we discuss how to efficiently construct the semidefinite programs that implement testing for sos in the Bell algebra $\mathbb{B}_2$.

Each nc monomial in $\mathbb{R}\langle \underline{X}, \underline{Y}\rangle$ with $m+n$ variables and degree d permits a unique numerical representation as a d-dimensional vector with elements from the set $\{1, 2, \dots, m+n\}$. The components are obtained as follows. To each X_i, $i = 1, \dots, m$, we assign its index i, while to each Y_j, $j = 1, \dots, n$, we assign $m+j$. For example if $m = 2$ and $n = 3$, then the monomial $X_1X_2Y_1Y_3$ has the vector representation $[1, 2, 3, 5]^T$. The constant monomial is represented by zero. Using this representation each polynomial is then given by the vector of coefficients and the associated vectors determining individual monomials.

Example 3.6. The polynomial $p = 1 - 2X + 2X^2 - XY - YX + Y^2 \in \mathbb{R}\langle X, Y\rangle$ is numerically represented as

$$\text{coef} = [1, -2, 2, -1, -1, 1]^T,$$
$$p_1 = 0, \quad p_2 = 1, \quad p_3 = [1, 1]^T, \quad p_4 = [1, 2]^T, \quad p_5 = [2, 1]^T, \quad p_6 = [2, 2]^T.$$

Furthermore, to each monomial in $\mathbb{R}\langle \underline{X}, \underline{Y}\rangle$ with $m+n$ variables and degree d we can assign a nonnegative integer, representing the position of this monomial in the graded lexicographic order. This can be seen by using a bijective correspondence between nonnegative integers and the vectors representing the monomials, as described above. This bijection is as follows:

- Vector 0 (representing constant monomial 1) is mapped to 1.
- Every vector $[a_{d-1}, \dots, a_0]^T$ representing a non-constant monomial is mapped to the integer

$$a_{d-1}(m+n)^{d-1} + a_{d-1}(m+n)^{d-2} + \ldots + a_1(m+n) + a_0 + 1. \tag{3.8}$$

- Conversely, we can assign to each integer $1 < p \leq ((m+n)^{d+1} - 1)/(m+n-1)$ the vector $[a_{d-1}, \dots, a_0]^T$, where

$$\begin{aligned} a_0 &= p - 1 - q_0(m+n), \qquad & q_0 &= \left\lceil \frac{p-1}{m+n} \right\rceil - 1, \\ a_1 &= q_0 - q_1(m+n), & q_1 &= \left\lceil \frac{q_0}{m+n} \right\rceil - 1, \\ &\vdots \\ a_{d-1} &= q_{d-2}. \end{aligned}$$

In case $p = 1$, we assign to it the constant monomial 1, i.e., the vector 0 in our framework.

Example 3.7. For $m = 1$, $n = 1$ listing all the nc monomials from $\mathbb{R}\langle \underline{X}, \underline{Y}\rangle$ up to degree $d = 3$ in graded lexicographic order

$$1,\ X,\ Y,\ X^2,\ XY,\ YX,\ Y^2,\ X^3,\ X^2Y,\ XYX,\ XY^2,\ YX^2,\ YXY,\ Y^2X,\ Y^3,$$

we see that the index of YXY is 13. This can also be computed from (3.8). The monomial YXY is represented as $[2,1,2]^T$, hence we obtain the index as

$$[2,1,2]\cdot[4,2,1]^T+1=13.$$

Let d_f be the degree of $f\in\mathbb{B}_2$ and let d denote the level of hierarchy, i.e., we consider the d-th order relaxation for $\lambda_{\min}(f)$, cf. Corollary 3.3:

$$\lambda_d := \sup\{\lambda \mid f-\lambda\in\Sigma^2_{2d}\}.$$

Recall that $f\in\mathbb{B}_2$ is a sum of degree d squares iff there exists a positive semidefinite matrix G satisfying $\mathbb{V}_d^\star G\mathbb{V}_d=f$, where $\mathbb{V}_d$ is the vector of all words in $\mathbb{B}_2$ of degree $\le d$.

Before we set up the linear equations for the semidefinite program (3.4) that connect various elements of matrix G, we apply a preprocessing step in which we prepare a table of reduced monomials from $\mathbb{B}_2$. More specifically, for a given hierarchy level d we first generate all the nc monomials from $\mathbb{R}\langle \underline{X},\underline{Y}\rangle$ up to degree d, and then we apply the commutative and idempotent relations from (3.1) to reduce every monomial so that no element is repeated and all the X_i are listed before any Y_j.

From (3.4) we see that the matrix E_{11} determines the objective function of the semidefinite program, while the equality constraints

$$f-f_1=\langle G-\langle E_{11},G\rangle E_{11},\mathbb{V}_d\mathbb{V}_d^\star\rangle \tag{3.9}$$

are obtained as follows: Among all monomials in $\mathbb{B}_2$, consider the ones having degree at most $d_{\max}=\lceil\frac{d}{2}\rceil$. They belong to $\mathbb{V}_d$. For every product $v_i^\star v_j$ between two words from $\mathbb{V}_d$, we find its reduced form and associated index in the generated table of all monomials up to degree d. The index determines the row in the system of linear equations that connects all the elements of the form $v_i^\star v_j$. If the product is in f, the associated component of the right hand side vector b equals its coefficient, otherwise it is zero. We solve the obtained semidefinite program

$$\inf\{\langle E_{11},G\rangle \mid A\mathrm{vec}(G)=b,\ G\succeq 0\}$$

with MOSEK [MOS19]. Since standard solvers fail to solve the $d=4$ relaxation for the Bell inequalities considered here, we present a regularization method in the next subsection to solve these as well. Finally, the value of the d-th order relaxation is then given by $\lambda_d=f_1-G_{1,1}$.

The procedure is demonstrated in the following simple example.

Example 3.8. Let us consider $f(X,Y)=X_1+Y_1-X_1Y_1-X_1Y_2-Y_1X_2+X_2Y_2$. This is the nc polynomial of degree $d_f=2$ that is associated with the CHSH inequality (1.1). We are looking for its smallest eigenvalue. Suppose we want to compute λ_1, i.e., the first level of the hierarchy (3.3). In order to set up the linear equations, we need to consider all reduced monomials in $\mathbb{B}_2$ in four variables X_1, X_2, Y_1 and Y_2 of degree ≤ 2:

$$1,X_1,X_2,Y_1,Y_2,X_1X_2,X_2X_1,X_1Y_1,X_1Y_2,X_2Y_1,X_2Y_2,Y_1Y_2,Y_2Y_1.$$

Among these, the monomials of degree ≤ 1 define the set $\mathbb{V}_1$. Computing $\mathbb{V}_1\mathbb{V}_1^\star$ and comparing the entries with monomials in f, we obtain the following system of 10 linear equations:

$$\begin{array}{rlrl}
X_1: & G_{1,2}+G_{2,1}+G_{2,2}=1 & X_1Y_1: & G_{2,4}+G_{4,2}=-1\\
X_2: & G_{1,3}+G_{3,1}+G_{3,3}=0 & X_1Y_2: & G_{2,5}+G_{5,2}=-1\\
Y_1: & G_{1,4}+G_{4,1}+G_{4,4}=1 & X_2Y_1: & G_{3,4}+G_{4,3}=-1\\
Y_2: & G_{1,5}+G_{5,1}+G_{5,5}=0 & X_2Y_2: & G_{3,5}+G_{5,3}=1\\
X_1X_2: & G_{2,3}+G_{3,2}=0 & Y_1Y_2: & G_{4,5}+G_{5,4}=0
\end{array}$$

Note that each non-symmetric monomial contributes to the same equation as its involution, and the equation connecting the constant monomial is omitted due to (3.9). The value of the relaxation is -0.2071068.

3.4. **Solving large semidefinite programs via regularization.** The hierarchy of semidefinite programs produces a sequence of lower bounds λ_d that converges to the optimal value $\lambda_{\min}$ of the problem (3.2). For instances for which the value of the current level does not match the optima given in the literature, we increase d and compute the next of level of the hierarchy. However, the resulting SDP on the fourth level optimizes a linear function over matrices of size 1486 or 2276 and has 249,315 or 568,675 linear equations in the case when $m+n=9$ or $m+n=10$, respectively. It is clear that the number of constraints will be a challenge for an interior-point solver. We tried to apply the Splitting Conic Solver (SCS) [OCPB16], a first-order method that scales to very large problems, at the cost of lower accuracy. However, we use the GNS construction (see Section 4) to extract the minimizers from the solution of the dual problem. For this to succeed we need to solve the SDP to higher precision.

To overcome these difficulties we propose to use regularization. The idea of approximating the original (linear) semidefinite program with a sequence of SDPs having a quadratic objective function is a known technique for solving large-scale SDPs [KMR17, HP21, HP22, ZST10, YST15]. The primal-dual pair of problems (3.4) and (3.5) can we written in standard form as

$$\begin{aligned} \sup \quad & \langle E_{11}, X\rangle \\ \text{s.t.} \quad & \mathcal{A}(X) = b \\ & X \succeq 0, \end{aligned} \tag{3.10}$$

$$\begin{aligned} \inf \quad & b^T y \\ \text{s.t.} \quad & \mathcal{A}^T(y) - E_{11} = Z \\ & Z \succeq 0, \end{aligned} \tag{3.11}$$

for an appropriate operator $\mathcal{A}\colon \mathbb{S}_d \to \mathbb{R}^p$, where p is the number of linear equations. We propose to use the augmented Lagrangian method to solve (3.11). We introduce a Lagrange multiplier X for the dual equation and consider for a penalty parameter $\alpha > 0$ the augmented Lagrangian function $\mathcal{L}_\alpha$:

$$\begin{aligned} \mathcal{L}_\alpha(y, Z; X) &= b^T y + \left\langle X, E_{11} - \mathcal{A}^T(y) + Z \right\rangle + \frac{1}{2\alpha} \left\| E_{11} - \mathcal{A}^T(y) + Z \right\|_F^2 \\ &= b^T y + \frac{1}{2\alpha} \left\| E_{11} - \mathcal{A}^T(y) + Z + \alpha X \right\|_F^2 - \frac{\alpha}{2} \|X\|_F^2 . \end{aligned}$$

This is the usual Lagrangian with an additional redundant quadratic term. The augmented Lagrangian method to solve (3.11) now consists in minimizing $\mathcal{L}_\alpha(y, Z; X)$ to get y and $Z \succeq 0$. Then the primal matrix X is updated $X \leftarrow X + \frac{1}{\alpha}\left(E_{11} - \mathcal{A}^T(y) + Z\right)$, see [Ber14, Section 2.2]. Then as $\alpha \to 0$ the whole processes is iterated until convergence.

The crucial part is how we solve the inner minimization problem

$$\inf\ b^T y + \frac{1}{2\alpha} \left\| E_{11} - \mathcal{A}^T(y) + Z + \alpha X \right\|_F^2 \quad \text{such that} \quad y \text{ free}, \ Z \succeq 0. \tag{3.12}$$

One way is to use the alternating optimization technique to minimize the dual function first with respect to y and then with respect to the dual variable Z. This leads to the famous alternating method of multipliers (ADMM) and its special variant called the Boundary Point Method proposed in [PRW06] in the context of semidefinite programming. As with SCS, this solver also lacks the ability to generate highly accurate solutions. Moreover, in the y update step, we need to solve a system of linear equations whose left-hand side has the form $\mathcal{A}\left(\mathcal{A}^T(y)\right)$. Considering the large number of constraints in our case, this is too costly.

We propose the following. The inner minimization problem (3.12) can be further simplified by eliminating Z as follows. Define $M = E_{11} - \mathcal{A}^T(y) + \alpha X$. Then for fixed y the problem

$$\inf_{Z \succeq 0} \|Z + M\|_F^2$$

is a projection of $-M$ onto the cone of positive semidefinite matrices. It is known that the solution $Z = (-M)_+$ can be computed from the eigenvalue decomposition of M, see [Hig88]. More specifically, if the eigenvalue decomposition of M is given by $X = S\mathrm{Diag}(\lambda)S^T$ with the eigenvalues $\lambda \in \mathbb{R}^k$ and orthogonal matrix $S \in \mathbb{R}^{k\times k}$, then we have $X_+ = S\mathrm{Diag}(\lambda_+)S^T$.

By substituting Z into (3.12) we obtain

$$\inf\ b^T y + \frac{1}{2\alpha} \left\| \left(E_{11} - \mathcal{A}^T(y) + \alpha X\right)_+ \right\|_F^2 \text{ such that } y \in \mathbb{R}^p.$$

Note that the objective function is convex and differentiable with explicit expressions of its gradient. In particular, if we denote the objective function as $F_\alpha(y)$, then we have

$$\nabla_y F_\alpha(y) = b - \frac{1}{\alpha}\mathcal{A}\left((E_{11} - \mathcal{A}^T(y) + \alpha X)_+\right).$$

The function value and the gradient are evaluated by computing the partial spectral decomposition. Then for a fixed α and X, the function is minimized using the L-BFGS algorithm [LN89]. Finally, by using $Z = (-M)_+$ the update on X is given by

$$X \leftarrow \left(X + \frac{1}{\alpha}\left(E_{11} - \mathcal{A}^T(y)\right)\right)_+.$$

The penalty parameter α regulates the tightness level of the solutions. We start with $\alpha = 10$ and after each iteration we decrease its value by 0.9 until it is sufficiently small.

3.5. **Reducing the size of the Gram matrix using an SOS conditional expectation.** Once again, we switch to viewing $\mathbb{B}_2$ as the group algebra $\mathbb{R}[\mathfrak{G}] = \mathbb{R}[\mathfrak{G}_m \times \mathfrak{G}_n] \cong \mathbb{R}[\mathfrak{G}_m] \otimes \mathbb{R}[\mathfrak{G}_n]$ as in the proof of Proposition 3.4. Let $f = \sum a_g g \in \mathbb{R}[\mathfrak{G}]$, and denote

$$\operatorname{supp} f := \{g \in \mathfrak{G} \mid a_g \neq 0\}.$$

We let $\mathfrak{G}_f$ be the subgroup of $\mathfrak{G}$ generated by $\operatorname{supp} f$.

Proposition 3.9. *For $f \in \mathbb{R}[\mathfrak{G}]$ the following are equivalent:*

(i) *f is a sum of squares in $\mathbb{R}[\mathfrak{G}]$, i.e., there are $h_j \in \mathbb{R}[\mathfrak{G}]$ such that*

$$f = \sum_j h_j^* h_j; \tag{3.13}$$

(ii) *f is a sum of squares in $\mathbb{R}[\mathfrak{G}_f]$, i.e., there are $h_j \in \mathbb{R}[\mathfrak{G}_f]$ such that (3.13) holds.*

Proof. The implication (ii) $\Rightarrow$ (i) is obvious. For the converse, consider the mapping

$$\mathbb{E} : \mathbb{R}[\mathfrak{G}] \to \mathbb{R}[\mathfrak{G}_f]$$
$$\mathfrak{G} \ni g \mapsto \begin{cases} g, & g \in \mathfrak{G}_f \\ 0, & \text{otherwise.} \end{cases}$$

This map is an SOS conditional expectation [SS13, Section 3], i.e., $\mathbb{E}$ is a unital $\mathbb{R}[\mathfrak{G}_f]$-module map, $\mathbb{E}(f)^* = \mathbb{E}(f^*)$ for all $f \in \mathbb{R}[\mathfrak{G}]$, and $\mathbb{E}$ maps sums of squares (in $\mathbb{R}[\mathfrak{G}]$) to sums of squares (in $\mathbb{R}[\mathfrak{G}_f]$). (See also [BHK] for a recent application of this notion to quantum games.) Thus applying $\mathbb{E}$ to a sum of squares of the form (3.13) leads to another sum of squares expression of the form (3.13) where the participating terms h_j are all in $\mathbb{R}[\mathfrak{G}_f]$, as desired. ☐

Proposition 3.9 suggests an alternate relaxation scheme for optimizing a Bell polynomial $f \in \mathbb{B}_2 = \mathbb{R}[\mathfrak{G}]$. Let $\mathcal{V}_1 = \overrightarrow{\operatorname{supp} f}$ be the vector having the support of f as its entries (including 1 even if $1 \notin \operatorname{supp} f$).[3] For $r \in \mathbb{N}$, let $\mathcal{V}_r$ be the vector of all distinct words of degree $\leq r$ in the entries of $\mathcal{V}_1$.

Lemma 3.10. *$f \in \mathbb{B}_2$ is a sum of squares iff there is $r \in \mathbb{N}$ $G \succeq 0$ such that*

$$f = \mathcal{V}_r^* G \mathcal{V}_r.$$

Proof. Immediate from Proposition 3.9. ☐

We can now pose an alternate SDP hierarchy to the optimization problem (3.2) for $f \in \mathbb{B}_2 = \mathbb{R}[\mathfrak{G}]$. For $d \in \mathbb{N}$ consider

$$\begin{array}{ll} \sup & f_1 - \langle E_{11}, G\rangle \\ \text{s.t.} & f - f_1 \equiv \mathcal{V}_d^\star \left(G - \langle E_{11}, G\rangle E_{11}\right)\mathcal{V}_d \\ & G \succeq 0, \end{array} \tag{3.14}$$

[3] An alternate might require adding all the variables to $\operatorname{supp} f$ leading to larger but tighter relaxations.

where f_1 denotes the constant term of f and E_{11} is the matrix unit with a one in the $(1,1)$ entry. By Lemma 3.10 and Proposition 3.1, the optimal values $\mu_d(f)$ of (3.14) increase with d and converge to $\lambda_{\min}(f)$ as $d \to \infty$.

As in Section 3.2, the dual SDP to (3.14) can be presented as

$$\begin{array}{rl} \inf & \langle G_f, H\rangle \\ \text{s.t.} & H_{1,1} = 1 \\ & H_{u,v} = H_{p,q}, \text{ for all } u,v,p,q \in \mathcal{V}_d : u^\star v \equiv p^\star q \\ & H \succeq 0, \end{array} \tag{3.15}$$

where the matrix G_f is a symmetric Gram matrix for f w.r.t. $\mathcal{V}_d$, i.e., $f = \mathcal{V}_d^* G_f \mathcal{V}_d$.

Example 3.11. Recall the CHSH polynomial

$$f = X_1 + Y_1 - X_1Y_1 - X_1Y_2 - X_2Y_1 + X_2Y_2 \in \mathbb{B}_2$$

from Example 3.8 which can in the Δ, Γ coordinates of Proposition 3.4 be written as

$$f = \frac{1}{2} - \frac{1}{4}(\Delta_1\Gamma_1 + \Delta_1\Gamma_2 + \Delta_2\Gamma_1 - \Delta_2\Gamma_2) \in \mathbb{R}[\mathfrak{G}].$$

Solving the primal-dual pair of SDPs (3.14), (3.15) with $d = 1$, where

$$\mathcal{V}_1 = \begin{pmatrix} 1 & \Delta_1\Gamma_1 & \Delta_1\Gamma_2 & \Delta_2\Gamma_1 & \Delta_2\Gamma_2 \end{pmatrix}^T, \qquad G_f = \begin{pmatrix} \frac{1}{2} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & \frac{1}{8} \\ -\frac{1}{8} & 0 & 0 & 0 & 0 \\ -\frac{1}{8} & 0 & 0 & 0 & 0 \\ -\frac{1}{8} & 0 & 0 & 0 & 0 \\ \frac{1}{8} & 0 & 0 & 0 & 0 \end{pmatrix},$$

leads to $\mu_1(f) = \frac{1}{2}\left(1 - \sqrt{2}\right)$,

$$G = \begin{pmatrix} \frac{1}{2\sqrt{2}} & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} & \frac{1}{8} \\ -\frac{1}{8} & \frac{1}{8\sqrt{2}} & \frac{1}{16\sqrt{2}} & \frac{1}{16\sqrt{2}} & 0 \\ -\frac{1}{8} & \frac{1}{16\sqrt{2}} & \frac{1}{8\sqrt{2}} & 0 & -\frac{1}{16\sqrt{2}} \\ -\frac{1}{8} & \frac{1}{16\sqrt{2}} & 0 & \frac{1}{8\sqrt{2}} & -\frac{1}{16\sqrt{2}} \\ \frac{1}{8} & 0 & -\frac{1}{16\sqrt{2}} & -\frac{1}{16\sqrt{2}} & \frac{1}{8\sqrt{2}} \end{pmatrix}, \qquad H = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 & 0 & 0 & -1 \\ \frac{1}{\sqrt{2}} & 0 & 1 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 1 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & -1 & 0 & 0 & 1 \end{pmatrix}$$

whence by Cholesky or eigenvalue decomposition we extract

$$f - \mu_1(f) = \frac{1}{16\sqrt{2}}\Big((2\sqrt{2} - \Delta_1\Gamma_1 - \Delta_1\Gamma_2 - \Delta_2\Gamma_1 + \Delta_2\Gamma_2)^2 + (\Delta_1\Gamma_1 + \Delta_2\Gamma_2)^2 + (-\Delta_1\Gamma_2 + \Delta_2\Gamma_1)^2\Big).$$

Example 3.12. For a slightly more complicated example, consider the A16 polynomial given in group algebra form as

$$f = \frac{1}{4}\Big(7 + 2\Delta_1 + \Delta_2 + \Delta_3 - \Delta_4 + 2\Gamma_1 + \Gamma_2 + \Gamma_3 - \Gamma_4 + \Delta_1\Gamma_1 + \Delta_1\Gamma_2 + \Delta_1\Gamma_3 - \Delta_1\Gamma_4 + \Delta_2\Gamma_1 - \Delta_2\Gamma_2 + \Delta_2\Gamma_3 + \Delta_3\Gamma_1 - \Delta_3\Gamma_3 - \Delta_3\Gamma_4 - \Delta_4\Gamma_1 - \Delta_4\Gamma_2 - \Delta_4\Gamma_4\Big)$$

In this case the border vector $\mathcal{V}_1$ has 22 entries, and $\mu_1(f) = -0,347072$. However, already the second relaxation (leading to an SDP of size 241×241) is exact, $\mu_2(f) = -0,300364$. Namely, the dual solution H and its restriction to the top 22×22 block have rank 16, which makes it in this case possible to extract optimizers thus certifying optimality, cf. Section 4.

On the other hand, without using the SOS conditional expectations, i.e., testing optima using the border vectors $\mathbb{V}_d$ as explained above in Section 3.2, required going to the third relaxation (of size 318×318) to obtain the exact value from the SDP, and the fourth relaxation (of size 1486×1486) was required to obtain flatness, extract optimizers and certify optimality (cf. Section 6).

To keep consistency with existing results in the physics literature, our numerical experiments in Section 6 are performed in the X, Y coordinates of the Bell algebra $\mathbb{B}_2$ and do not make use of the SOS conditional expectations.

4. Extracting minimizers and certifying optimality

In this section we describe how the well-known flatness condition for the solution of the dual SDP (3.5) can be exploited to extract minimizers, thus certifying optimality, and even improving accuracy in certain situations.

4.1. Flat extensions and the GNS construction.

In this subsection we show how to extract optimizers under the flatness assumption on the solution to the dual SDP (3.5).

Given a matrix Z that is feasible for (3.5), the objective function, the so-called Riesz functional

$$L : f \mapsto \langle G_f, Z\rangle$$

is a well-defined linear functional $(\mathbb{B}_2)_{2d} \to \mathbb{R}$ that is positive in the sense that it is nonnegative on every sum of degree $\leq d$ squares in $\mathbb{B}_2$, i.e., $L(\Sigma^2_{2d}) \succeq \mathbb{R}_{\geq 0}$.

We call L or the associated Hankel matrix $Z_L = Z$ *flat* if the rank of Z_L is the same as the rank of $Z_{\check{L}}$, where $\check{L}$ denotes the restriction of L to $(\mathbb{B}_2)_{2d-2}$.

Lemma 4.1. *A linear functional $L : (\mathbb{B}_2)_{2d} \to \mathbb{R}$ is positive iff the associated Hankel matrix Z_L is positive semidefinite.*

Proof. For $p = \sum_w p_w w \in (\mathbb{B}_2)_{2d}$ let $\mathbf{p}$ denote the vector of its coefficients so that $p = \mathbf{p}^* \mathbb{V}_d$. Then for every $p, q \in (\mathbb{B}_2)_d$ we have

$$L(p^\star q) = \sum_{u,v} p_u q_v L(u^\star v) = \sum_{u,v} p_u q_v (Z_L)_{u,v} = \mathbf{p}^* Z_L \mathbf{q}$$

yielding the desired conclusion. □

Theorem 4.2. *Suppose the optimizer H of* (3.5) *is flat. Then with $r = \operatorname{rank} H$ there exist tuples of $r \times r$ symmetric matrices $\underline{x}$, $\underline{y}$ and a unit vector v satisfying* (3.1) *and*

$$\langle G_f, H\rangle = \langle f(\underline{x}, \underline{y})v, v\rangle,$$

i.e., the objective value of (3.5)*, which is equal to λ_d by Corollary 3.5, is equal to the global minimum $\lambda_{\min}(f)$.*

Proof. We construct the tuples using a Gelfand-Naimark-Segal (GNS) construction. Since the nc Hankel matrix H is positive semidefinite, we can find a Gram decomposition $H = [\langle \mathbf{u}, \mathbf{w}\rangle]_{u,w}$ with vectors $\mathbf{u}, \mathbf{w} \in \mathbb{R}^r$, where the labels are words in $\mathbb{B}_2$ of degree at most d. Using this decomposition we set

$$\mathcal{H} = \operatorname{span}\{\mathbf{w} \mid \deg w \leq d\}.$$

By the flatness assumption one gets that

$$\mathcal{H} = \operatorname{span}\{\mathbf{w} \mid \deg w \leq d\} = \operatorname{span}\{\mathbf{w} \mid \deg w \leq d-1\}. \tag{4.1}$$

The Riesz functional L associated to H defines an inner product on $\mathcal{H}$ via

$$(\mathbf{p}, \mathbf{q}) \mapsto L(q^\star p),$$

thus $\mathcal{H}$ is a finite dimensional Hilbert space. Hence we can consider the left regular representations, i.e., the operators x_i represent the left multiplication by X_i on $\mathcal{H}$, and likewise for y_i. That is, $x_i \mathbf{w} = \mathbf{X_i w}$ and $y_i \mathbf{w} = \mathbf{Y_i w}$. Since by equation (4.1) we only need to consider words w with $\deg w \leq d-1$, the resulting words $X_i w$ and $Y_i w$ are of degree $\leq d$. Hence the $x_i, y_i : \mathcal{H} \to \mathcal{H}$ are well-defined.

We claim that x_i, y_i are symmetric. For all $p, q \in (\mathbb{B}_2)_{d-1}$ we have

$$(x_j \mathbf{p}, \mathbf{q}) = L(q^\star (X_j p)) = L((X_j q)^\star p) = (\mathbf{p}, x_j \mathbf{q}),$$

whence $x_i^* = x_i$. A similar calculation establishes $y_j^* = y_j$.

We next show that $\underline{x}, \underline{y}$ satisfy (3.1). Given $p, q \in (\mathbb{B}_2)_{d-1}$,

$$(x_j^2 \mathbf{p}, \mathbf{q}) = (x_j \mathbf{p}, x_j \mathbf{q}) = L((X_j q)^\star X_j p) = L(q^\star X_j^2 p) = L(q^\star X_j p) = (x_j \mathbf{p}, \mathbf{q}),$$

whence $x_j^2 = x_j$. By symmetry, $y_i^2 = y_i$. Finally,

$$\begin{aligned}(x_i y_j \mathbf{p}, \mathbf{q}) &= (y_j \mathbf{p}, x_i \mathbf{q}) = L((X_i q)^\star Y_j p) = L(q^\star X_i Y_j p) = L(q^\star Y_j X_i p) = L((Y_j q)^\star X_i p)\\ &= (x_i \mathbf{p}, y_j \mathbf{q}) = (y_j x_i \mathbf{p}, \mathbf{q}),\end{aligned}$$

establishing (3.1).

Let $\mathbf{1} \in \mathcal{H}$ be the vector corresponding to the constant element $1 \in \mathbb{B}_2$. Then

$$\|\mathbf{1}\|^2 = (\mathbf{1}, \mathbf{1}) = L(1^\star 1) = L(1) = 1.$$

Finally,

$$\lambda_{\min}(f) \le (f(\underline{x}, \underline{y})\mathbf{1}, \mathbf{1}) = L(f) \le \lambda_{\min}(f),$$

concluding the proof. □

Corollary 4.3. *Suppose the optimizer H of (3.5) is flat, and let $r = \operatorname{rank} H$. There are $r_1, r_2 \in \mathbb{N}$ with $r = r_1 r_2$, and there exist tuples $\underline{x}$ of $r_1 \times r_1$ hermitian matrices, $\underline{y}$ of $r_2 \times r_2$ hermitian matrices, and a unit vector v satisfying (3.1) and*

$$\langle G_f, H\rangle = \langle f(\underline{x} \otimes I_{r_2}, I_{r_1} \otimes \underline{y})v, v\rangle = \lambda_{\min}(f). \tag{4.2}$$

Proof. This is a straightforward consequence of Theorem 4.2. Indeed, start with the tuples $\underline{x}$, $\underline{y}$ produced by Theorem 4.2 and then apply Proposition 2.2 to obtain (up to unitary equivalence) the desired tensor decomposition as in (4.2) for $\underline{x}, \underline{y}$. □

If flatness is observed, then optimizers can be extracted using the Gelfand-Naimark-Segal (GNS) construction. Observe that the dual matrix

$$H = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

is flat iff $B = AZ$ and $C = Z^T AZ$ for some matrix Z. In practice flatness is often accompanied with numerical issues. It is thus worthwhile to measure deviation from flatness. We express H's deviation from flatness by computing

$$\mathrm{err}_{\mathrm{flat}} = \frac{\|C - Z^T AZ\|_F}{1 + \|C\|_F + \|Z^T AZ\|_F} \tag{4.3}$$

using the Frobenius norm. Observe that being flat is independent of the choice of Z, but that $\mathrm{err}_{\mathrm{flat}}$ does depend on Z. (We refer the reader to Figure 1 below for a demonstration of soundness of this definition.) After extracting the submatrices A and B from H, we obtain Z by solving the matrix equation $B = AZ$ in MATLAB as $Z = A\backslash B$. If $\mathrm{err}_{\mathrm{flat}}$ is nonzero but very small, one can still attempt to use the GNS construction to obtain approximate solutions [KPV18] to (3.1). Sometimes these can be rounded to obtain accurate approximate solutions.

4.2. **Improving accuracy: rounding based on the Artin-Wedderburn theory.** In theory computing a tensor decomposition as in Proposition 2.2 for tuples $\underline{x}, \underline{y}$ satisfying (3.1) is routine. But in practice the tuples obtained, say from the GNS construction from flatness via Theorem 4.2 or from approximate flatness as in [KPV18] only satisfy (3.1) approximately. In this case we can project onto the form of Proposition 2.2 to increase exactness and precision. The resulting matrices $x_j = \xi_j \otimes I_{r_2}$ and $y_k = I_{r_1} \otimes \gamma_k$ then satisfy the commutative relations by construction, since

$$x_j y_k = (\xi_j \otimes I_{r_2})(I_{r_1} \otimes \gamma_k) = (I_{r_1} \otimes \gamma_k)(\xi_j \otimes I_{r_2}) = y_k x_j,$$

while idempotence is achieved by rounding the eigenvalues of the ξ_j, γ_k to 0 and 1. We describe all this in some more detail next.

In general, solutions to the optimization problem (3.5) tend to be linear functionals $L \in (\sigma_{2d}^2)^\vee$ that are extreme points. If the corresponding matrix H is flat of rank r, then the obtained GNS construction will yield an irreducible tuple [Arv76, Theorem 1.6.6].[4] That is, the subalgebra of $M_r(\mathbb{C})$ generated by the x_j, y_k is $M_r(\mathbb{C})$ itself. Letting $\mathcal{X}$ denote the subalgebra generated by the x_j, and likewise for $\mathcal{Y}$, we have

$$M_r(\mathbb{C}) \cong \mathcal{X} \otimes \mathcal{Y}$$

[4]In the sequel we restrict to tuples of complex matrices; a similar though slightly more involved analysis works over the reals.

since $\mathcal{X}, \mathcal{Y}$ commute. Thus the tuples $\underline{x}$ and $\underline{y}$ are both irreducible as well. That is, $\mathcal{X} \cong M_{r_1}(\mathbb{C})$ and $\mathcal{Y} \cong M_{r_2}(\mathbb{C})$ with $r = r_1 r_2$. Since the only embeddings $M_{r_1}(\mathbb{C}) \to M_{r_1 r_2}(\mathbb{C})$ are of the form $A \mapsto U^*(A \otimes I_{r_2})U$ for some $r \times r$ unitary, the matrices x_j have to be (after a unitary change of bases) of the form $\xi_j \otimes I_{r_2}$. From the commuting relations in (3.1) and irreducibility we deduce the y_k are of the form $I_{r_1} \otimes \gamma_k$ then.

The first step of the post-processing procedure to improve the accuracy of the minimizers consists of running a numerical algorithm (e.g. [MKKK10]) for the Artin-Wedderburn block-diagonal decomposition of matrix $*$-algebras [Lam13] described above. It produces an orthogonal matrix Q such that all the matrices $Q^T y_j Q$ are simultaneously block diagonalized, i.e., the matrices Y_j are decomposed into a direct sum as

$$Q^T y_j Q = \bigoplus_i y_j^{(i)}.$$

For instances that we consider and for which the optimizers can be extracted from flatness, it turned out that the rank of the dual matrix is either 4, 8, 9, 16 or 25. For the cases when the rank r is a perfect square $r = r_1^2$ we can find an orthogonal matrix Q for which the tuples of matrices $\underline{x}$ and $\underline{y}$ are approximately of the form of Corollary 4.3, i.e.,

$$\underline{x} \approx \underline{\xi} \otimes I_{r_1}, \quad \underline{y} \approx I_{r_1} \otimes \underline{\gamma} \tag{4.4}$$

for some tuples of matrices $\underline{\xi}$ and $\underline{\gamma}$. If the norm of the difference between the left hand side and the right hand size in equation (4.4) is less than the threshold 10^{-5}, we replace $\underline{x}$ and $\underline{y}$ by the respective right-hand side of (4.4). This brings the matrices into the required form. Finally, note that the two tensor products appearing in (4.4) commute automatically.

The situation is similar when the rank is 8. In this case the ξ_j, γ_k will be 2×2 *complex* self-adjoint matrices, and we first identify their 4×4 real versions with the same process described above. Here, an $n \times n$ complex matrix A corresponds to the real $2n \times 2n$ matrix by

$$A = B + iC \leftrightsquigarrow \begin{bmatrix} B & C \\ -C & B \end{bmatrix}.$$

This also covers the case when the rank is 8 and the underlying matrix is real. Then $C = 0$ and

$$A = \begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$$

is the direct product of two rank 4 matrices and we proceed as before.

Finally, by computing the spectral decomposition of each of the matrices ξ_j and γ_k obtained from (4.4) and rounding the eigenvalues to 0 and 1, the matrices $x_i = \xi_i \otimes I$, $i = 1, \dots, m$ and $y_j = I \otimes \gamma_j$, $j = 1, \dots, n$ satisfy the relations from (3.1).

4.3. Certifying non-optimality of a dual SDP solution arising from the hierarchy using noncommutative Gröbner bases.

The solution H to the dual SDP (3.5) for the d-th relaxation yields a positive linear functional

$$L \in (\Sigma_{2d}^2)^\vee \text{ with } L(1) = 1. \tag{4.5}$$

In the presence of the flat condition (cf. Section 4.1) such an L is of the form

$$L(f) = \langle f(\underline{x} \otimes I, I \otimes \underline{y})v, v\rangle, \qquad f \in (\mathbb{B}_2)_{2d} \tag{4.6}$$

for some unit vector v and tuples of orthogonal projections $\underline{x}, \underline{y}$. In this case L is called a (truncated) Bell moment functional. If L is defined on the entire $\mathbb{B}_2$ and (4.6) holds for all $f \in \mathbb{B}_2$ without degree restriction, then L is called a full Bell moment functional.

The main result of this section presents a necessary condition for Bell moment functionals. In particular, this may help certify that a given L is *not* a Bell moment functional, i.e., L cannot be written in the form (4.6). In the commutative case the analog theory was developed by Curto and Fialkow in their solution of the truncated moment problem [CF98]. They call the condition recursive generation (RG).

Proposition 4.4. *Let $L \in (\Sigma^2_{2d})^\vee$ with $L(1) = 1$. Consider the associated Hankel matrix $H = H_L$ and form the left ideal $\mathfrak{a} \subseteq \mathbb{B}_2$ generated by $\ker H$.*[5] *If L is a Bell moment functional, then $L(\mathfrak{a} \cap (\mathbb{B}_2)_{2d}) = \{0\}$.*

Proof. Suppose L is a Bell moment functional, i.e., (4.6) holds. Then the formula in (4.6) extends L to a full Bell moment functional, denoted $\hat{L} : \mathbb{B}_2 \to \mathbb{R}$. The positive semidefinite (infinite) Hankel matrix $H_{\hat{L}}$ extends H_L in the sense that it is of the form

$$H_{\hat{L}} = \begin{pmatrix} H_L & \star \\ \star & \star \end{pmatrix}.$$

In particular, $\mathcal{N} := \ker H_L \subseteq \ker H_{\hat{L}} =: \hat{\mathcal{N}}$. For any $b \in \hat{\mathcal{N}}$, the associated polynomial $p_b \in \mathbb{B}_2$ satisfies $\hat{L}(p_b) = 0$ and

$$\hat{L}(wp_b) = 0$$

for any $e \in \mathbb{N}$ and $w \in \mathbb{V}_e$. Thus $\hat{L}$ must vanish on all the elements in the left ideal generated by $\mathcal{N}$. In particular, the same must be true for L if one restricts to elements of degree $\leq 2d$. $\square$

Proposition 4.4 gives rise to Algorithm 1 that depends on noncommutative Gröbner bases theory. Loosely speaking, a Gröbner basis (GB) of an ideal is a particularly well-behaved set of generators for the ideal. For instance, it enables us to answer the ideal membership problem. We refer the reader to [Mor86, Gre00, MR98, Xiu12] for details.

Algorithm 1: ncRG algorithm:
a necessary condition for $L \in (\Sigma^2_{2d})^\vee$ to be a Bell moment functional

---

Input : $L \in (\Sigma^2_{2d})^\vee$ with $L(1) = 1$
Output: Does L satisfy the ncRG condition?

Step 1: Compute a basis b for the kernel of the Hankel matrix associated to L. Let B be a lift of b to $\mathbb{R}\langle \underline{X}, \underline{Y} \rangle$, i.e., simply replace all lower case x, y with upper case X, Y
Step 2: Compute a noncommutative Gröbner basis G for the ideal $\mathfrak{A}$ generated by $\{sZ \mid s \in B\} \cup \{X_i^2 - X_i, Y_j^2 - Y_j, [X_i, Y_j] \mid i, j\}$ in $\mathbb{R}\langle \underline{X}, \underline{Y}, Z \rangle$
Step 3: Find a basis B_0 for $\{f \in (\mathbb{B}_2)_{2d} \mid FZ \in \mathfrak{A}\}$; here F is the lift of f to an element of $\mathbb{R}\langle \underline{X}, \underline{Y} \rangle$
Step 4: Output whether $L|_{B_0} = 0$

Remark 4.5. To justify the algorithm, we need to check the noncommutative RG (ncRG) condition from Proposition 4.4. [BHK, Proposition 6.1] shows that for $f \in \mathbb{B}_2$, $f \in \mathfrak{a}$ iff $FZ \in \mathfrak{A}$ (in the notation of Algorithm 1). Thus $L|_{B_0} = 0$ iff L satisfies $L(\mathfrak{a} \cap \mathbb{B}_{2d}) = \{0\}$, as desired.

Example 4.6. We give a demonstration of how Algorithm 1 can be applied. Consider $m = n = 2$. Let $L : (\mathbb{B}_2)_4 \to \mathbb{R}$ be determined by its Hankel matrix H whose columns are

$$\left(\begin{matrix} 1 & v_1 & v_2 \\ v_1 & v_1 & v_5 \\ v_2 & v_5 & v_2 \\ \frac{1}{2}(-v_4 + v_5 + v_{11} - v_{14} + 2) & \frac{1}{2}(-v_4 + v_5 - 2v_9 - v_{11} + v_{14} + 2) & v_9 \\ v_4 & 2v_1 + v_4 + 2v_9 + v_{11} - 2v_{14} - 2 & v_{10} \\ v_5 & v_5 & -2v_2 + 2v_9 + v_{10} + v_{14} \\ v_5 & v_{13} & v_5 \\ \frac{1}{2}(-v_4 + v_5 - 2v_9 - v_{11} + v_{14} + 2) & \frac{1}{2}(-v_4 + v_5 - 2v_9 - v_{11} + v_{14} + 2) & v_{14} \\ 2v_1 + v_4 + 2v_9 + v_{11} - 2v_{14} - 2 & 2v_1 + v_4 + 2v_9 + v_{11} - 2v_{14} - 2 & \frac{1}{2}(-2v_2 - v_4 + 2v_5 + 2v_9 + v_{10} + v_{11} - 2v_{14}) \\ v_9 & v_{14} & v_9 \\ v_{10} & \frac{1}{2}(-2v_2 - v_4 + 2v_5 + 2v_9 + v_{10} + v_{11} - 2v_{14}) & v_{10} \\ v_{11} & 0 & \frac{1}{2}(-2v_2 + v_4 + 2v_5 + 2v_9 + v_{10} - v_{11} - 2v_{14}) \\ v_{11} & 0 & \frac{1}{2}(-2v_2 + v_4 + 2v_5 + 2v_9 + v_{10} - v_{11} - 2v_{14}) \end{matrix} \right.$$

[5] This is a slight abuse of notation. The columns of the matrix H are indexed by words $\mathbb{V}_d$ in the Bell algebra of degree $\leq d$. Thus a nullvector for H naturally represents an element of $(\mathbb{B}_2)_d$.

$$
\begin{matrix}
\frac{1}{2}\left(-v_4+v_5+v_{11}-v_{14}+2\right) & v_4 & v_5 \\
\frac{1}{2}\left(-v_4+v_5-2v_9-v_{11}+v_{14}+2\right) & 2v_1+v_4+2v_9+v_{11}-2v_{14}-2 & v_5 \\
v_9 & v_{10} & -2v_2+2v_9+v_{10}+v_{14} \\
\frac{1}{2}\left(-v_4+v_5+v_{11}-v_{14}+2\right) & v_{11} & v_{14} \\
v_{11} & v_4 & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) \\
v_{14} & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) & -2v_2+2v_9+v_{10}+v_{14} \\
v_{14} & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) & v_{25} \\
\frac{1}{2}\left(-v_4+v_5-2v_9-v_{11}+v_{14}+2\right) & 0 & v_{14} \\
0 & 2v_1+v_4+2v_9+v_{11}-2v_{14}-2 & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) \\
v_9 & \frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) & \frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}\right) \\
\frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) & v_{10} & -2v_2+v_4+2v_5+2v_9-v_{11}-2v_{14}+v_{28} \\
v_{11} & 2v_1+2v_9+3v_{11}-2v_{14}-2 & \frac{1}{2}\left(4v_1-2v_2+v_4+2v_5+6v_9+v_{10}+3v_{11}-6v_{14}-4\right) \\
v_{23} & v_{11} & v_{28}
\end{matrix}
$$

$$
\begin{matrix}
v_5 & \frac{1}{2}\left(-v_4+v_5-2v_9-v_{11}+v_{14}+2\right) & 2v_1+v_4+2v_9+v_{11}-2v_{14}-2 \\
v_{13} & \frac{1}{2}\left(-v_4+v_5-2v_9-v_{11}+v_{14}+2\right) & 2v_1+v_4+2v_9+v_{11}-2v_{14}-2 \\
v_5 & v_{14} & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) \\
v_{14} & \frac{1}{2}\left(-v_4+v_5-2v_9-v_{11}+v_{14}+2\right) & 0 \\
\frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) & 0 & 2v_1+v_4+2v_9+v_{11}-2v_{14}-2 \\
v_{25} & v_{14} & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) \\
v_{13} & v_2+\frac{v_4}{2}+v_5-v_9-\frac{v_{10}}{2}-\frac{v_{11}}{2}-v_{14}+v_{25} & v_{27} \\
v_2+\frac{v_4}{2}+v_5-v_9-\frac{v_{10}}{2}-\frac{v_{11}}{2}-v_{14}+v_{25} & \frac{1}{2}\left(-v_4+v_5-2v_9-v_{11}+v_{14}+2\right) & 0 \\
v_{27} & 0 & 2v_1+v_4+2v_9+v_{11}-2v_{14}-2 \\
v_{14} & v_{14} & \frac{1}{2}\left(4v_1-2v_2+v_4+2v_5+6v_9+v_{10}+3v_{11}-6v_{14}-4\right) \\
\frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) & v_{28} & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) \\
v_{28} & 0 & v_{31} \\
\frac{1}{2}\left(4v_1-2v_2+v_4+2v_5+6v_9+v_{10}+3v_{11}-6v_{14}-4\right) & v_{30} & 0
\end{matrix}
$$

$$
\begin{matrix}
v_9 & v_{10} \\
v_{14} & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) \\
v_9 & v_{10} \\
v_9 & \frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) \\
\frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) & v_{10} \\
\frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}\right) & -2v_2+v_4+2v_5+2v_9-v_{11}-2v_{14}+v_{28} \\
v_{14} & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) \\
v_{14} & v_{28} \\
\frac{1}{2}\left(4v_1-2v_2+v_4+2v_5+6v_9+v_{10}+3v_{11}-6v_{14}-4\right) & \frac{1}{2}\left(-2v_2-v_4+2v_5+2v_9+v_{10}+v_{11}-2v_{14}\right) \\
v_9 & \frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) \\
\frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) & v_{10} \\
\frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) & v_{38} \\
v_{11}-v_{23}+v_{28}-v_{30} & \frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right)
\end{matrix}
$$

$$
\left.
\begin{matrix}
v_{11} & v_{11} \\
0 & 0 \\
\frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) & \frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) \\
v_{11} & v_{23} \\
2v_1+2v_9+3v_{11}-2v_{14}-2 & v_{11} \\
\frac{1}{2}\left(4v_1-2v_2+v_4+2v_5+6v_9+v_{10}+3v_{11}-6v_{14}-4\right) & v_{28} \\
v_{28} & \frac{1}{2}\left(4v_1-2v_2+v_4+2v_5+6v_9+v_{10}+3v_{11}-6v_{14}-4\right) \\
0 & v_{30} \\
v_{31} & 0 \\
\frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) & v_{11}-v_{23}+v_{28}-v_{30} \\
v_{38} & \frac{1}{2}\left(-2v_2+v_4+2v_5+2v_9+v_{10}-v_{11}-2v_{14}\right) \\
2v_1+2v_9+3v_{11}-2v_{14}-2 & v_{39} \\
v_{39} & v_{23}
\end{matrix}
\right),
$$

where

$$
v_1=\frac{19}{79},\ v_2=\frac{23}{44},\ v_4=\frac{31}{66},\ v_5=\frac{18}{125},\ v_9=\frac{450}{901},\ v_{10}=\frac{7}{41},\ v_{11}=\frac{25}{72},
$$
$$
v_{13}=\frac{15}{128},\ v_{14}=\frac{9}{61},\ v_{23}=\frac{27}{101},\ v_{25}=\frac{4}{33},\ v_{27}=-\frac{1}{1976},\ v_{28}=\frac{1}{178},
$$
$$
v_{30}=0,\ v_{31}=\frac{1}{1119},\ v_{38}=\frac{3}{52},\ v_{39}=\frac{8}{43},
$$

and whose columns are indexed by

$$
(1,\ x_1,\ x_2,\ y_1,\ y_2,\ x_1x_2,\ x_2x_1,\ x_1y_1,\ x_1y_2,\ x_2y_1,\ x_2y_2,\ y_1y_2,\ y_2y_1).
$$

Thus, for instance, $L(x_1x_2y_1)=\frac{9}{61}$.

Then $H \succeq 0$ has one-dimensional kernel spanned by

$$
\begin{pmatrix} -2 & 0 & 0 & 1 & 1 & -1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}^T.
$$

This vector represents the element $-2+y_1+y_2-x_1x_2+x_1y_1+x_2y_1 \in \mathbb{B}_2$.

By Step 2 of Algorithm 1 we are thus led to find the noncommutative Gröbner basis G for the ideal $\mathfrak{A} \subseteq \mathbb{R}\langle X_1, X_2, Y_1, Y_2, Z\rangle$ generated by

$$
\{(-2+Y_1+Y_2-X_1X_2+X_1Y_1+X_2Y_1)Z,
$$

$$X_1^2 - X_1,\ X_2^2 - X_2,\ Y_1^2 - Y_1,\ Y_2^2 - Y_2,\ [X_1, Y_1],\ [X_2, Y_1],\ [X_1, Y_2],\ [X_2, Y_2]\}.$$

While running a GB algorithm with NCGB in NCAlgebra under Mathematica or Magma it appears that G will be infinite. Producing a finite truncation leads to the following set of polynomials:

$$\begin{gathered}\{-X_1 + X_1^2,\ \ -X_2 + X_2^2,\ \ -X_1Y_1 + Y_1X_1,\ \ -X_2Y_1 + Y_1X_2,\ \ -Y_1 + Y_1^2,\ \ -X_1Y_2 + Y_2X_1,\\ -X_2Y_2 + Y_2X_2,\ \ -Y_2 + Y_2^2,\ \ -2Z + Y_1Z + Y_2Z - X_1X_2Z + X_1Y_1Z + X_2Y_1Z,\\ 2Z - 2X_1Z - 2Y_1Z - Y_2Z + 2X_1Y_1Z + X_1Y_2Z + Y_1Y_2Z,\\ 4Z - 2X_2Z - 2Y_1Z - 2Y_2Z + 2X_1X_2Z - 2X_1Y_1Z + X_2Y_2Z - X_2X_1X_2Z + X_2X_1Y_1Z,\\ -Y_2Z + Y_2Y_1Z - X_1X_2Y_2Z + X_1Y_2Y_1Z + X_2Y_2Y_1Z,\\ 2Y_2Z - X_2Y_2Z - 2Y_2Y_1Z + 2X_1X_2Y_2Z - 2X_1Y_2Y_1Z - X_2X_1X_2Y_2Z + X_2X_1Y_2Y_1Z,\\ 2Z - 14X_1Z - 2Y_1Z - Y_2Z + 14X_1Y_1Z + 7X_1Y_2Z - 2X_1X_2X_1Z - 3X_1X_2Y_2Z + Y_1Y_2Y_1Z + 2X_1X_2X_1X_2Z + X_1X_2X_1Y_2Z + X_1Y_1Y_2Y_1Z + X_2Y_1Y_2Y_1Z\}.\end{gathered}$$

Now

$$\begin{aligned}F = 1 - X_2 + X_1X_2 - X_1Y_1 + \frac{1}{2}X_2X_1 + \frac{1}{4}X_2Y_2 - \frac{1}{2}Y_1Y_2 - \frac{1}{2}Y_2Y_1 + \frac{1}{4}X_1X_2Y_2\\ - \frac{1}{2}X_1Y_2Y_1 - \frac{1}{4}X_2X_1X_2 - \frac{1}{6}X_2X_1Y_2 + \frac{1}{4}X_1X_2X_1X_2 - \frac{1}{4}X_1X_2X_1Y_1\\ - \frac{1}{4}X_2X_1X_2Y_1 - \frac{1}{6}X_2X_1X_2Y_2 - \frac{1}{12}X_2Y_2Y_1Y_2\end{aligned}$$

satisfies $FZ \in \mathfrak{A}$ as can be checked by computing its canonical form modulo the (truncated) Gröbner basis. However,

$$L(f) = -\frac{131149699493}{563151950712000} \neq 0,$$

where f denotes the image of F in $\mathbb{B}_2$. Thus by Algorithm 1, L is not ncRG and thus cannot be a Bell moment functional.

5. Upper bounds on violations of Bell's inequalities via nonlinear optimization

The approach based on the hierarchy of semidefinite programs produces a sequence of lower bounds λ_d that converges to the optimal value $\lambda_{\min}$ of the problem (3.2). For many instances we can extract the minimizers by computing the appropriate level of the hierarchy and using Gelfand-Naimark-Segal (GNS) construction, as explained in Section 4. For instances for which the value of the current level of the hierarchy matches the optima given in the literature but the flatness condition is not met, we would need to compute the next of level of the hierarchy. For some instances even the SDP relaxation on the fourth level was not enough to certify optimality and extract minimizers.

In order to prove optimality for such instances, we propose to use nonlinear optimization. We explicitly formulate (3.2) as a non-convex optimization problem

$$\begin{aligned} &\inf \quad \left\langle f(\underline{X}, \underline{Y}), vv^T \right\rangle \\ &\text{s.t.} \quad \underline{X}, \underline{Y} \text{ satisfy (3.1)},\ \|v\| = 1 \end{aligned} \tag{5.1}$$

and use standard nonlinear optimization methods to compute its local minima. This produces multiple upper bounds on the violation of a Bell inequality. In the event that there is no gap between the lower bound obtained via the SDP approach and an upper bound obtained from the local minimum, we can certify that the matrices computed via (5.1) are indeed optimal.

Using Corollary 4.3 we anticipate that if minimizers $\underline{X}, \underline{Y}$ of size $r \times r$ exist, we search for the optimal tuples $\underline{X}$ and $\underline{Y}$ of the form

$$\underline{X} = \underline{x} \otimes I_{r_1}, \quad \underline{Y} = I_{r_2} \otimes \underline{y}, \tag{5.2}$$

for some r_1, $r_2 \in \mathbb{N}$ such that $r = r_1 r_2$.

We use properties of the matrix tuples $\underline{x}$ and $\underline{y}$ to reduce the number of variables that are needed to parametrize them. Note that these matrices are symmetric and idempotent. The only non-singular idempotent matrix is the identity matrix. Furthermore, the zero matrix is also feasible for problem (5.1). Apart from these, we need to consider matrices with rank at least 1. It turns out that for most of the cases we consider, rank-1 matrices are sufficient, i.e., the minima of (5.1) are attained at matrices $\underline{X}$ and $\underline{Y}$ for which $\underline{x}$ and $\underline{y}$ from (5.2) are either zero matrices, identity matrices or have rank exactly 1. In the following we list how rank-1 matrices can be parametrized depending on the dimension. By using the properties

$$A = A^H, \quad \operatorname{rank}(A) = 1 \quad \text{and} \quad A^2 = A \tag{5.3}$$

and using the fact that the trace of an idempotent matrix equals its rank, we get:

- $\mathbb{R}^2$: If
$$A = \begin{bmatrix} a & b \\ b & 1-a \end{bmatrix}$$
the condition $A^2 = A$ reduces to the quadratic equation
$$a^2 - a + b^2 = 0 \quad \text{or} \quad \left(a - \frac{1}{2}\right)^2 + b^2 = \frac{1}{4},$$
which is a circle with center $\left(\frac{1}{2}, 0\right)$ and radius $\frac{1}{2}$. This gives the parametrization
$$A(u) = \frac{1}{2}\begin{bmatrix} 1-\cos u & \sin u \\ \sin u & 1+\cos u \end{bmatrix}.$$
We can also use the rational parametrization
$$A(x) = \frac{1}{1+x^2}\begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix}$$
that can easily be generalized to higher dimensions. Hence each of the matrices in (5.2) can be defined by 1 variable if we set
$$X_i = A(\theta_i) \otimes I_2, \quad Y_j = I_2 \otimes A(\lambda_j)$$
for some θ_i, $i = 1, \ldots, m$ and λ_j, $j = 1, \ldots, n$. Together with the eigenvector $v \in \mathbb{R}^4$ this gives $n + m + 4$ decision variables in the constrained optimization problem (5.1), if $f \in \mathbb{B}_2$ was given with m variables $\underline{X}$ and n variables $\underline{Y}$.

- $\mathbb{R}^3$: The conditions (5.3) give
$$A(x,y) = \frac{1}{1+x^2+y^2}\begin{bmatrix} 1 & x & y \\ x & x^2 & xy \\ y & xy & y^2 \end{bmatrix}$$
In this case the matrices in (5.2) can be parametrized with 2 variables. Together with the eigenvector $v \in \mathbb{R}^9$ this yields $2(m+n)+9$ decision variables.

- $\mathbb{C}^2$: If
$$A = \begin{bmatrix} a & \alpha + i\beta \\ \alpha - i\beta & 1-a \end{bmatrix}$$
the condition $A^2 = A$ reduces to
$$a^2 - a + \alpha^2 + \beta^2 = 0 \quad \text{or} \quad \left(a - \frac{1}{2}\right)^2 + \alpha^2 + \beta^2 = \frac{1}{4},$$
which is a sphere with center $\left(\frac{1}{2}, 0, 0\right)$ and radius $\frac{1}{2}$. This gives the parametrization
$$A(u,v) = \frac{1}{2}\begin{bmatrix} 1+\cos u \sin v & \sin u \sin v + i\cos v \\ \sin u \sin v - i \cos v & 1 - \cos u \sin v \end{bmatrix}.$$
Similarly to the real case we can also use the parametrization
$$A(x,y) = \frac{1}{1+x^2+y^2}\begin{bmatrix} 1 & x+iy \\ x-iy & x^2+y^2 \end{bmatrix}.$$

Again the matrices in (5.2) can thus be described with 2 variables. Since the eigenvector lies in $\mathbb{C}^4$, this gives $2(m+n)+8$ decision variables in total.

We minimize the objective function from (5.1) where the feasible matrices are constrained to being either zero, identity or rank-1. Hence there are 3^{m+n} possible optimization problems in total, if $f \in \mathbb{B}_2$ has $m+n$ variables. In all cases we used MATLAB's built in function `fmincon` with multiple random starts to compute the local optimums. The results are summarized in Table 3. It is remarkable that optimality for so many cases can be proved using this strategy.

6. Numerical results

6.1. **Data.** In this section we report results obtained on the 88 instances of Bell inequalities labeled as A2 to A89. These instances were used in the computational experiments described in [IIA06]. The article [AIIS05] explains how these Bell inequalities are obtained. We have rewritten them in a noncommutative polynomial optimization formulation. More precisely, for each instance we created a txt file, containing the numbers of variables X_i and Y_i, i.e., m and n, the polynomial f for which we seek to compute the minimum eigenvalue to obtain the largest violation and all the relations that apply to the variables. These data are available at `https://github.com/HrgaT/Bell_inequality_data`.

6.2. **Results.** We organize the results into four groups.

(1) The first group of results, reported in Table 1, contains results on instances, for which the optimum value from the literature was achieved by the SDP hierarchy for some d, the dual solution was flat and the minimizer was extracted by the GNS construction, and the optimum value was additionally certified by the nonlinear programming approach. We want to point out three instances, A38, A52 and A81, for which we corrected the violation of the underlying Bell inequality by 10^{-7}.
(2) The second group of results, reported in Table 2, contains the results where the optimum was achieved by the SDP hierarchy for some d, the dual solution was flat and the minimizer was extracted by the GNS construction, but nonlinear programming approach did not give the optimum value.
(3) The third group of results, reported in Table 3, contains the results where the optimum from the literature was attained by the SDP hierarchy for some d, but the minimizer was not reconstructed, even when going all the way to the level four hierarchy. However, the nonlinear programming approach delivered the optimum value which serves as a certificate of optimality.
(4) The fourth group of instances, reported in Table 4, contains the rest of the results and is split into two parts. (a) Instances for which we could recompute the optimum obtained from literature [PV09, PV10], but could not confirm the optimality either by GNS or by the nonlinear programming approach. Nevertheless, the bounds from the literature together with the bounds from the SDP hierarchy certify the optimality for these five instances (up to an accuracy of 10^{-8}). New here is the confirmation of the maximal violation for A80, where by computing the level 3 of the SDP hierarchy we managed to close the gap between the upper and lower bound; (b) Remaining instances. Here the finite sequence of lower bounds λ_d is increasing but does not reach the best known upper bounds from the literature [PV09, PV10]. For all these instances we solved the level four SDP hierarchy. For A14, A21, A47, A62, A64, A68, A82, A84 and A89 we have thus improved the currently best known lower bound on the violation. Notice that by going from level 3 to level 4 the SDP bound for A82 stays the same.

In the following tables we report the optimum violation from the literature [PV09, PV10] (column opt), the optimum value of the SDP hierarchy (λ_d) attained at level d. The difference between both optima is in the column $|\text{opt} - \lambda_d|$. When we could extract the optimizer by the GNS construction, we report d_{GNS}, the level of the dual SDP hierarchy for which the optimum matrix was flat. For these cases we also report the rank of these flat solutions (column rank) and the numerical deviations from flatness err_{flat} introduced in (4.3). We report err_{flat} also for the instances where the GNS construction did not give the optimum matrices (Tables 3–4) to demonstrate that in these cases this parameter is significantly larger compared to the cases where the GNS construction was successful.

For the instances in Tables 1 and 3 we confirmed the optimum (also) with the NLP approach, therefore we provide in these two tables also the optimum values $\mathrm{opt}_{\mathrm{NLP}}$ and the differences $|\mathrm{opt} - \mathrm{opt}_{\mathrm{NLP}}|$.

TABLE 1. Numerical results on instances, where the optimum value from the literature (opt) was achieved by the SDP hierarchy in level d, reported in column 4, the minimizer was extracted by GNS on level d_{GNS} from the dual optimum of rank, reported in column 7. The optimum value was confirmed by NLP, i.e., the value $\mathrm{opt}_{\mathrm{NLP}}$ in column 9 is equal to opt and λ_d is within a numerical error of 10^{-8} (10^{-7} for A38, A52, A81).

Instance	opt	λ_d	d	$\lvert\mathrm{opt} - \lambda_d\rvert$	d_{GNS}	rank	$\mathrm{err}_{\mathrm{flat}}$	$\mathrm{opt}_{\mathrm{NLP}}$	$\lvert\mathrm{opt} - \mathrm{opt}_{\mathrm{NLP}}\rvert$
A2	-0.2071068	-0.2071068	1	0.0000000	3	4	0.0000000	-0.2071068	0.0000000
A4	-0.2990381	-0.2990381	3	0.0000000	3	4	0.0000000	-0.2990381	0.0000000
A5	-0.4353342	-0.4353342	3	0.0000000	3	4	0.0000000	-0.4353342	0.0000000
A9	-0.4652428	-0.4652428	3	0.0000000	3	8	0.0000000	-0.4652428	0.0000000
A11	-0.4561079	-0.4561079	3	0.0000000	3	8	0.0000000	-0.4561079	0.0000000
A12	-0.4877093	-0.4877093	2	0.0000000	3	8	0.0000000	-0.4877093	0.0000000
A15	-0.4496279	-0.4496279	3	0.0000000	3	8	0.0000000	-0.4496279	0.0000000
A16	-0.4571068	-0.4571068	3	0.0000000	4	8	0.0000000	-0.4571068	0.0000000
A17	-0.3754473	-0.3754473	2	0.0000000	2	4	0.0000000	-0.3754473	0.0000000
A22	-0.6234571	-0.6234571	3	0.0000000	3	8	0.0000001	-0.6234571	0.0000000
A23	-0.5460735	-0.5460735	2	0.0000000	3	8	0.0000000	-0.5460735	0.0000000
A24	-0.6047986	-0.6047986	2	0.0000000	2	4	0.0000000	-0.6047986	0.0000000
A25	-0.6033789	-0.6033789	2	0.0000000	2	4	0.0000000	-0.6033789	0.0000000
A27	-0.6483073	-0.6483073	2	0.0000000	2	4	0.0000000	-0.6483073	0.0000000
A28	-0.6403143	-0.6403143	2	0.0000000	3	4	0.0000000	-0.6403143	0.0000000
A29	-0.4920635	-0.4920635	2	0.0000000	3	8	0.0000000	-0.4920635	0.0000000
A30	-0.5698209	-0.5698209	2	0.0000000	2	4	0.0000000	-0.5698209	0.0000000
A31	-0.5738173	-0.5738173	2	0.0000000	2	4	0.0000000	-0.5738173	0.0000000
A32	-0.4135530	-0.4135530	3	0.0000000	3	8	0.0000000	-0.4135530	0.0000000
A33	-0.6226313	-0.6226313	3	0.0000000	3	8	0.0000000	-0.6226313	0.0000000
A34	-0.5350117	-0.5350117	3	0.0000000	3	4	0.0000000	-0.5350117	0.0000000
A35	-0.6249079	-0.6249079	3	0.0000000	3	4	0.0000000	-0.6249079	0.0000000
A36	-0.4388685	-0.4388685	3	0.0000000	3	8	0.0000000	-0.4388685	0.0000000
A37	-0.4868868	-0.4868868	3	0.0000000	3	8	0.0000000	-0.4868868	0.0000000
A38	-0.4699126	-0.4699127	3	0.0000001	3	8	0.0000000	-0.4699127	0.0000001
A39	-0.6172035	-0.6172035	2	0.0000000	3	8	0.0000000	-0.6172035	0.0000000
A40	-0.6078638	-0.6078638	2	0.0000000	2	4	0.0000000	-0.6078638	0.0000000
A41	-0.4785634	-0.4785634	3	0.0000000	3	8	0.0000000	-0.4785634	0.0000000
A43	-0.6107654	-0.6107654	2	0.0000000	2	4	0.0000000	-0.6107654	0.0000000
A44	-0.5364942	-0.5364942	3	0.0000000	3	4	0.0000000	-0.5364942	0.0000000
A45	-0.5372394	-0.5372394	3	0.0000000	3	8	0.0000000	-0.5372394	0.0000000
A49	-0.4666943	-0.4666943	3	0.0000000	3	8	0.0000000	-0.4666943	0.0000000
A50	-0.5182900	-0.5182900	3	0.0000000	3	8	0.0000000	-0.5182900	0.0000000
A52	-0.6218611	-0.6218612	3	0.0000001	3	8	0.0000000	-0.6218612	0.0000001
A53	-0.6386102	-0.6386102	3	0.0000000	3	8	0.0000000	-0.6386102	0.0000000
A54	-0.5936813	-0.5936813	3	0.0000000	3	8	0.0000000	-0.5936813	0.0000000
A55	-0.6213203	-0.6213203	2	0.0000000	3	4	0.0000005	-0.6213203	0.0000000
A57	-0.6603444	-0.6603444	3	0.0000000	3	8	0.0000000	-0.6603444	0.0000000
A58	-0.6488905	-0.6488905	3	0.0000000	3	4	0.0000000	-0.6488905	0.0000000
A59	-0.4488256	-0.4488256	3	0.0000000	3	4	0.0000000	-0.4488256	0.0000000
A61	-0.4019248	-0.4019248	3	0.0000000	3	8	0.0000000	-0.4019248	0.0000000
A66	-0.4877093	-0.4877093	3	0.0000000	3	8	0.0000000	-0.4877093	0.0000000
A70	-0.6052228	-0.6052228	2	0.0000000	3	8	0.0000000	-0.6052228	0.0000000
A71	-0.4490163	-0.4490163	3	0.0000000	3	8	0.0000000	-0.4490163	0.0000000
A72	-0.6962822	-0.6962822	3	0.0000000	3	4	0.0000000	-0.6962822	0.0000000
A73	-0.8831381	-0.8831381	3	0.0000000	3	4	0.0000000	-0.8831381	0.0000000
A74	-0.6890694	-0.6890694	3	0.0000000	3	8	0.0000000	-0.6890694	0.0000000
A75	-0.6051510	-0.6051510	3	0.0000000	3	4	0.0000000	-0.6051510	0.0000000
A76	-0.4898631	-0.4898631	3	0.0000000	3	9	0.0000000	-0.4898631	0.0000000
A77	-0.6655582	-0.6655582	3	0.0000000	3	4	0.0000000	-0.6655582	0.0000000
A78	-0.8927018	-0.8927018	2	0.0000000	2	4	0.0000000	-0.8927018	0.0000000
A79	-0.6243153	-0.6243153	3	0.0000000	3	8	0.0000000	-0.6243153	0.0000000

Continued on next page

Table 1 – *Continued from previous page*

Instance	opt	λ_d	d	$\lvert\text{opt} - \lambda_d\rvert$	d_{GNS}	rank	err_{flat}	opt_{NLP}	$\lvert\text{opt} - \text{opt}_{\text{NLP}}\rvert$
A81	-0.6690099	-0.6690100	3	0.0000001	3	8	0.0000000	-0.6690100	0.0000001
A83	-0.6961664	-0.6961664	3	0.0000000	3	8	0.0000000	-0.6961664	0.0000000
A85	-0.6411408	-0.6411408	3	0.0000000	3	8	0.0000000	-0.6411408	0.0000000
A86	-0.8004425	-0.8004425	3	0.0000000	3	8	0.0000000	-0.8004425	0.0000000

Table 2. Numerical results on instances where the optimum was achieved by the SDP hierarchy for some d, minimizer was extracted by GNS on level d_{GNS}, but the nonlinear programming approach did not produce the global optimum.

Instance	opt	λ_d	d	$\lvert\text{opt} - \lambda_d\rvert$	d_{GNS}	rank	err_{flat}
A6	-0.3003638	-0.3003638	3	0.0000000	3	16	0.0000000
A10	-0.4158004	-0.4158004	3	0.0000000	3	25	0.0000000
A13	-0.4252330	-0.4252330	3	0.0000000	3	16	0.0000000
A46	-0.4590108	-0.4590108	3	0.0000000	3	25	0.0000000
A48	-0.4631707	-0.4631707	3	0.0000000	3	16	0.0000000
A63	-0.4894164	-0.4894164	3	0.0000000	3	16	0.0000000
A65	-0.3688996	-0.3688996	3	0.0000000	3	25	0.0000000

Table 3. This table contains the results for instances where the optimum was achieved by the SDP hierarchy for some d, but the dual solution was not flat so we could not reconstruct the minimizer. However, the nonlinear programming bounds coincide with the SDP bounds, which confirms the bounds are the global optima.

Instance	opt	λ_d	d	$\lvert\text{opt} - \lambda_d\rvert$	err_{flat}	opt_{NLP}	$\lvert\text{opt} - \text{opt}_{\text{NLP}}\rvert$
A7	-0.2878683	-0.2878683	3	0.0000000	0.0023000	-0.2878683	0.0000000
A8	-0.5916501	-0.5916501	1	0.0000000	0.0200000	-0.5916501	0.0000000
A18	-0.3843551	-0.3843551	2	0.0000000	0.0165500	-0.3843551	0.0000000
A19	-0.6226300	-0.6226300	3	0.0000000	0.0200000	-0.6226300	0.0000000
A20	-0.6022398	-0.6022398	3	0.0000000	0.0045000	-0.6022398	0.0000000
A26	-0.5275550	-0.5275550	2	0.0000000	0.0053000	-0.5275555	0.0000005
A42	-0.6198655	-0.6198655	2	0.0000000	0.0100000	-0.6198655	0.0000000
A51	-0.6607809	-0.6607809	3	0.0000000	0.0144000	-0.6607809	0.0000000
A56	-0.6893124	-0.6893124	3	0.0000000	0.0059000	-0.6893124	0.0000000
A69	-0.6096103	-0.6096103	3	0.0000000	0.0200000	-0.6096103	0.0000000
A88	-0.4142136	-0.4142136	2	0.0000000	0.0190000	-0.4142136	0.0000000

Table 4. Numerical results for instances for which the SDP bounds are computed, but are not confirmed to be the optimum since the GNS did not yield minimizers and the optimum values, obtained by nonlinear programming, differ from the SDP bounds. The table is divided into two parts: the upper part contains instances where our SDP bounds are equal to the optima published elsewhere in the literature, while the lower part contains the rest of the instances, i.e., where we could not reach the bounds from the literature even with the level 4 hierarchy.

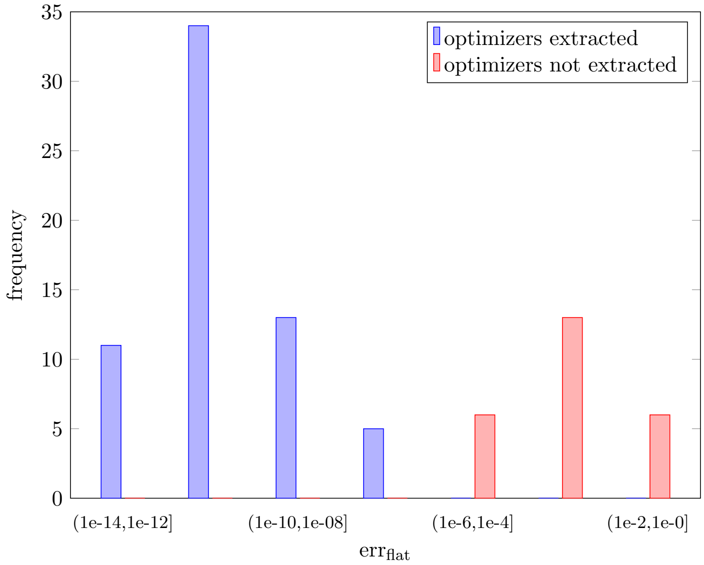
Instance	opt	λ_d	d	$\lvert\text{opt} - \lambda_d\rvert$	err_{flat}	λ_d	d	$\lvert\text{opt} - \lambda_d\rvert$	err_{flat}
A3	-0.2508754	-0.2508756	3	0.0000002	0.0082000	-0.2508754	4	0.0000000	0.0006800
A60	-0.3940032	-0.3940032	3	0.0000000	0.0004200	-0.3940032	4	0.0000000	0.0001000
A67	-0.3990671	-0.3990671	3	0.0000000	0.0003400	-0.3990671	4	0.0000000	0.0001000
A80	-0.3769863	-0.3769863	3	0.0000000	0.0010000	-0.3769863	4	0.0000000	0.0002100
A87	-0.7562471	-0.7562471	3	0.0000000	0.0130000	-0.7562471	4	0.0000000	0.0002000

Continued on next page

Table 4 – *Continued from previous page*

Instance	opt	λ_d	d	$\lvert\text{opt} - \lambda_d\rvert$	err_{flat}	λ_d	d	$\lvert\text{opt} - \lambda_d\rvert$	err_{flat}
A14	-0.4759513	-0.4778466	3	0.0018953	0.0100000	-0.4768133	4	0.0008620	0.0015000
A21	-0.3260601	-0.3261777	3	0.0001176	0.0085000	-0.3260654	4	0.0000053	0.0012000
A47	-0.4608544	-0.4616077	3	0.0007533	0.0038000	-0.4615622	4	0.0007078	0.0022000
A62	-0.4065268	-0.4067133	3	0.0001865	0.0011000	-0.4066728	4	0.0001460	0.0000039
A64	-0.3900890	-0.3906267	3	0.0005377	0.0004500	-0.3905074	4	0.0004184	0.0031000
A68	-0.4025522	-0.4050299	3	0.0024777	0.0008100	-0.4050294	4	0.0024772	0.0000070
A82	-0.4708838	-0.4709172	3	0.0000334	0.0016000	-0.4709172	4	0.0000334	0.0000072
A84	-0.6352087	-0.6352108	3	0.0000021	0.0004000	-0.6352107	4	0.0000020	0.0000012
A89	-0.3035637	-0.3056390	3	0.0020753	0.0069000	-0.3054510	4	0.0018873	0.0011000

FIGURE 1. Barplot depicting how the flatness error err_{flat} relates to the success of the GNS construction. Blue bars correspond to the instances where the minimizer was extracted by flatness (instances from Tables 1–2), while the red bars visualize the err_{flat} of the rest of the instances, where either (1) the optimum was attained but we could not extract the minimum by GNS (Table 3 or the first part of Table 4) or (2) the instances for which the computed SDP bounds are not confirmed to be the optimum (second half of Table 4).



Remark 6.1. Note that for all the instances from out data we can also exploit their sparsity pattern, as was done in [KMP22]. While this is a promising topic to be explored further, this approach leads to weaker lower bounds for at least some of the problems considered here. For example, for the problem A3 from Table 4, the sparse SDP hierarchy at level 3 gives the bound -0.2512, while the NPA hierarchy for this problem gives a much tighter optimum value already at level 2. Additionally, the implementation presented in the current paper is very efficient and we can compute the NPA bounds for level 4, while our implementation of the sparse SDP bounds from [KMP22] can reach only level 3.

7. Conclusion and Perspectives

In this paper we studied the hierarchy of SDP relaxations for the case of noncommutative polynomial optimization problems that arise from Bell inequalities. We work in the Bell algebra as the appropriate algebraic framework to study the violation of Bell inequalities. This leads to a reduction in size of the underlying SDP problems and paves the path towards computing optimum values of these hierarchies for levels up to 3 using standard SDP solvers, and for level 4 using our adaption of the regularization method, combined with the L-BFGS algorithm.

The main focus was certifying optimality of Bell inequality violations. We champion the traditional approach based on the flatness rank condition to extract optimizers via the so–called Gelfand-Naimark-Segal (GNS) construction, to certify optimality of a Bell inequality violation.

By applying Artin-Wedderburn theory to rounding and projecting we improved the accuracy of the obtained optimizers.

Further, we apply a standard nonlinear programming (NLP) formulation to obtain upper bounds on the optimum violation of Bell inequalities. Their importance are twofold: (i) when the GNS construction yielded an optimizer, NLP bound was a double confirmation for the optimum; (ii) if the optimum was correctly computed by the SDP bound but was not confirmed by the GNS construction, the NLP bound can be used to confirm the optimum value in case it coincides with the SDP bound.

We provided extensive numerical results on the list of 88 instances of Bell inequalities. On 79 of them the optimum value of the SDP hierarchy that we computed coincided with the optima from the literature and on 74 of them we could certify that these are indeed the optima: on 63 out of them the GNS construction yielded a certificate for optimality, and on the remaining 11 instances the NLP approach gave a numerical certificate of optimality.

Our paper therefore established a solid theoretical basis for certified optimization of noncommutative polynomials in the Bell algebra and also demonstrated that implementations of the algorithms that we proposed have high practical significance. One can naturally extend our approach to trace optimization, and to other algebras which are similar to the Bell algebra, like multipartite Bell algebras where the variables are partitioned into more than 2 classes of pairwise commuting tuples. Moreover, the potential for various notions of sparsity, which was only hinted at in this paper, needs to be explored further. Finally, many of the examples considered here exhibit natural symmetries which can lead to a reduction in size for the SDPs either through group actions on the SDP problem itself or by algebraic manipulations using invariant or semi-invariant noncommutative polynomials, which we also intend to pursue in further research.

References

[AIIS05] David Avis, Hiroshi Imai, Tsuyoshi Ito, and Yuuya Sasaki. Two-party Bell inequalities derived from combinatorics via triangular elimination. *J. Phys. A*, 38(50):10971, 2005. 2, 3, 21

[AM19] Amir Ali Ahmadi and Anirudha Majumdar. DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization. *SIAM J. Appl. Algebra Geom.*, 3(2):193–230, 2019. 1

[Arv76] William Arveson. *An invitation to C^*-algebras*. Graduate Texts in Mathematics, No. 39. Springer-Verlag, New York-Heidelberg, 1976. 15

[Bel64] John S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1(3):195, 1964. 1

[Ber14] Dimitri P Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014. 11

[BFS16] Mario Berta, Omar Fawzi, and Volkher B. Scholz. Quantum bilinear optimization. *SIAM J. Optim.*, 26(3):1529–1564, 2016. 2

[BHK] Adam Bene Watts, J. William Helton, and Igor Klep. Noncommutative Nullstellensätze and perfect games. *Ann. Henri Poincaré*. to appear. 12, 17

[BKP16] Sabine Burgdorf, Igor Klep, and Janez Povh. *Optimization of polynomials in non-commuting variables*. SpringerBriefs in Mathematics. Springer, [Cham], 2016. 1, 2, 3, 8

[BPT13] Grigoriy Blekherman, Pablo A. Parrilo, and Rekha R. Thomas, editors. *Semidefinite optimization and convex algebraic geometry*, volume 13 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2013. 3

[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. 4, 16

[CF00] Raul E. Curto and Lawrence A. Fialkow. The Truncated Complex K-Moment Problem. *Trans. Amer. Math. Soc.*, 352(6):pp. 2825–2855, 2000. 3

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden variable theories. *Phys. Rev. Lett.*, 23(15):880, 1969. 2

[CKP11] Kristijan Cafuta, Igor Klep, and Janez Povh. NCSOStools: a computer algebra system for symbolic and numerical computation with noncommutative polynomials. *Optim. Methods Softw.*, 26(3):363–380, 2011. 2

[CKP12] Kristijan Cafuta, Igor Klep, and Janez Povh. Constrained polynomial optimization problems with noncommuting variables. *SIAM J. Optim.*, 22(2):363–383, 2012. 1

[DLTW08] Andrew C. Doherty, Yeong-Cherng Liang, Ben Toner, and Stephanie Wehner. The quantum moment problem and bounds on entangled multi-prover games. In *Twenty-Third Annual IEEE Conference on Computational Complexity*, pages 199–210. IEEE Computer Soc., Los Alamitos, CA, 2008. 1, 2, 3, 6

[DOHMP09] Mauricio C. De Oliveira, J. William Helton, Scott A. Mccullough, and Mihai Putinar. *Engineering Systems and Free Semi-Algebraic Geometry*, pages 17–61. Springer New York, New York, NY, 2009. 2

[Fri12] Tobias Fritz. Tsirelson's problem and Kirchberg's conjecture. *Rev. Math. Phys.*, 24(5):1250012, 67, 2012. 7

[GdLL18] Sander Gribling, David de Laat, and Monique Laurent. Bounds on entanglement dimensions and quantum graph parameters via noncommutative polynomial optimization. *Math. Program.*, 170(1, Ser. B):5–42, 2018. 2

[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

[Gis09] Nicolas Gisin. Bell inequalities: many questions, a few answers. In *Quantum Reality, relativistic causality, and closing the epistemic circle*, pages 125–138. Springer, 2009. 2

[Gre00] Edward L. Green. Multiplicative bases, Gröbner bases, and right Gröbner bases. *J. Symbolic Comput.*, 29(4-5):601–623, 2000. 17

[Hel02] J. William Helton. "Positive" noncommutative polynomials are sums of squares. *Ann. of Math. (2)*, 156(2):675–694, 2002. 1, 2, 6

[Hig88] Nicholas J. Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl.*, 103:103–118, 1988. 11

[HLL09] Didier Henrion, Jean-Bernard Lasserre, and Johan Löfberg. GloptiPoly 3: moments, optimization and semidefinite programming. *Optim. Methods Softw.*, 24(4-5):761–779, 2009. 1

[HM04] J. William Helton and Scott A. McCullough. A Positivstellensatz for non-commutative polynomials. *Trans. Amer. Math. Soc.*, 356(9):3721–3737, 2004. 1, 3, 7

[HP21] Timotej Hrga and Janez Povh. MADAM: a parallel exact solver for Max-Cut based on semidefinite programming and ADMM. *Comput. Optim. Appl.*, 80(2):347–375, 2021. 11

[HP22] Timotej Hrga and Janez Povh. Solving SDP relaxations of max-cut problem with large number of hypermetric inequalities by L-BFGS-B. *Optimization Letters*, pages 1–13, 2022. 11

[IIA06] Tsuyoshi Ito, Hiroshi Imai, and David Avis. Bell inequalities stronger than the Clauser-Horne-Shimony-Holt inequality for three-level isotropic states. *Phys. Rev. A*, 73(4):042109, 2006. 3, 21

[JNP$^+$11] Marius Junge, Miguel Navascues, Carlos Palazuelos, David Perez-Garcia, Volkher B. Scholz, and Reinhard F. Werner. Connes embedding problem and Tsirelson's problem. *J. Math. Phys.*, 52(1):012102, 12, 2011. 7

[JNV$^+$21] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. MIP*=RE. *Commun. ACM*, 64(11):131–138, 2021. 7

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

[KMR17] Nathan Krislock, Jérôme Malick, and Frédéric Roupin. BiqCrunch: A semidefinite branch-and-bound method for solving binary quadratic problems. *ACM Trans. Math. Software (TOMS)*, 43(4):1–23, 2017. 11

[KPV18] Igor Klep, Janez Povh, and Jurij Volčič. Minimizer extraction in polynomial optimization is robust. *SIAM J. Optim.*, 28(4):3177–3207, 2018. 15

[Lam13] Tsit-Yuen Lam. *A first course in noncommutative rings*, volume 131. Springer Science & Business Media, 2013. 4, 16

[Las01] Jean-Bernard Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2000/01. 1

[Las10] Jean Bernard Lasserre. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series*. Imperial College Press, London, 2010. 1

[Lau09] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 157–270. Springer, New York, 2009. 1

[LN89] Dong C. Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Math. Program.*, 45(1):503–528, 1989. 12

[McC01] Scott McCullough. Factorization of operator-valued polynomials in several non-commuting variables. *Linear Algebra Appl.*, 326(1-3):193–203, 2001. 1, 3, 6

[MHL15] Victor Magron, Didier Henrion, and Jean-Bernard Lasserre. Semidefinite approximations of projections and polynomial images of semialgebraic sets. *SIAM J. Optim.*, 25(4):2143–2164, 2015. 1

[MKKK10] Kazuo Murota, Yoshihiro Kanno, Masakazu Kojima, and Sadayoshi Kojima. A numerical algorithm for block-diagonal decomposition of matrix $*$-algebras with application to semidefinite programming. *Jpn. J. Ind. Appl. Math.*, 27(1):125–160, 2010. 16

[Mor86] Ferdinando Mora. Groebner bases for noncommutative polynomial rings. In *Algebraic algorithms and error correcting codes (Grenoble, 1985)*, volume 229 of *Lecture Notes in Comput. Sci.*, pages 353–362. Springer, Berlin, 1986. 17

[MOS19] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019. 10

[MR98] Klaus Madlener and Birgit Reinert. String rewriting and Groebner bases – a general approach to monoid and group rings. *Technical report Zentrum Fur Computeralgebra Technische Universität Kaiserslautern*, page 55, May 1998. 17

[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

[Nie14] Jiawang Nie. Optimality conditions and finite convergence of Lasserre's hierarchy. *Math. Program.*, 146(1-2, Ser. A):97–121, 2014. 1

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New J. Phys.*, 10(7):073013, 2008. 1, 2, 3, 6, 7

[OCPB16] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *J. Optim. Theory Appl.*, 169(3):1042–1068, 2016. 11

[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. 1, 3

[PRW06] Janez Povh, Franz Rendl, and Angelika Wiegele. A boundary point method to solve semidefinite programs. *Computing*, 78:277–286, 2006. 11

[Put93] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.*, 42(3):969–984, 1993. 1

[PV09] Károly F. Pál and Tamás Vértesi. Quantum bounds on Bell inequalities. *Phys. Rev. A (3)*, 79(2):022120, 12, 2009. 2, 3, 21

[PV10] Károly F. Pál and Tamás Vértesi. Maximal violation of a bipartite three-setting, two-outcome Bell inequality using infinite-dimensional quantum systems. *Phys. Rev. A (3)*, 82(2):022116, 2010. 3, 21

[Rez78] Bruce Reznick. Extremal PSD forms with few terms. *Duke Math. J.*, 45(2):363–374, 1978. 3

[RFP10] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010. 5

[RKM+01] Mary A. Rowe, David Kielpinski, Volker Meyer, Charles A. Sackett, Wayne M. Itano, Christopher Monroe, and David J. Wineland. Experimental violation of a Bell's inequality with efficient detection. *Nature*, 409(6822):791–794, 2001. 2

[RTAL13] Cordian Riener, Thorsten Theobald, Lina Jansson Andrén, and Jean-Bernard Lasserre. Exploiting symmetries in SDP-relaxations for polynomial optimization. *Math. Oper. Res.*, 38(1):122–141, 2013. 1

[Sch05] Markus Schweighofer. Optimization of polynomials on compact semialgebraic sets. *SIAM J. Optim.*, 15(3):805–825, 2005. 1

[SS13] Yurii Savchuk and Konrad Schmüdgen. Unbounded induced representations of $*$-algebras. *Algebras Represent. Theory*, 16:309–376, 2013. 3, 12

[Tsi80] Boris S. Tsirelson. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.*, 4(2):93–100, 1980. 2

[Vid59] Ivan Vidav. On some *-regular rings. *Acad. Serbe Sci. Publ. Inst. Math.*, 13:73–80, 1959. 7

[WML21] Jie Wang, Victor Magron, and Jean-Bernard Lasserre. Chordal-TSSOS: a moment-SOS hierarchy that exploits term sparsity with chordal extension. *SIAM J. Optim.*, 31(1):114–141, 2021. 1

[Xiu12] Xingqiang Xiu. *Non-commutative Gröbner bases and applications*. PhD thesis, 2012. 17

[YST15] Liuqin Yang, Defeng Sun, and Kim-Chuan Toh. SDPNAL+: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Math. Program. Comput.*, 7(3):331–366, 2015. 11

[ZST10] Xin-Yuan Zhao, Defeng Sun, and Kim-Chuan Toh. A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM J. Optim.*, 20(4):1737–1765, 2010. 11

Igor Klep: Faculty of Mathematics and Physics, University of Ljubljana & Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia
Email address: igor.klep@fmf.uni-lj.si

Timotej Hrga: Faculty of Mathematics and Physics, University of Ljubljana, Slovenia
Email address: timotej.hrga@fmf.uni-lj.si

Janez Povh: Faculty of Mechanical Engineering, University of Ljubljana & Rudolfovo - Science and technology Center Novo Mesto, Slovenia
Email address: janez.povh@fs.uni-lj.si