Commun. Math. Phys. 366, 423–468 (2019)
Digital Object Identifier (DOI) https://doi.org/10.1007/s00220-019-03382-y

Communications in
Mathematical Physics

Limitations of Semidefinite Programs for Separable States and Entangled Games

Aram W. Harrow[1], **Anand Natarajan**[2], **Xiaodi Wu**[3]

[1] Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: anand@natarajans.org; anandn@caltech.edu
[3] Department of Computer Science, Institute for Advanced Computer Studies, and Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, MD 20742, USA

Received: 22 February 2017 / Accepted: 30 April 2018
Published online: 4 March 2019 – © Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract: Semidefinite programs (SDPs) are a framework for exact or approximate optimization with widespread application in quantum information theory. We introduce a new method for using reductions to construct integrality gaps for SDPs, meaning instances where the SDP value is far from the true optimum. These are based on new limitations on the sum-of-squares (SoS) hierarchy in approximating two particularly important sets in quantum information theory, where previously no $\omega(1)$-round integrality gaps were known:

1. The set of separable (i.e. unentangled) states, or equivalently, the $2 \to 4$ norm of a matrix.
2. The set of quantum correlations; i.e. conditional probability distributions achievable with local measurements on a shared entangled state.

Integrality gaps for the $2 \to 4$ norm had previously been sought due to its connection to Small-Set Expansion (SSE) and Unique Games (UG).

In both cases no-go theorems were previously known based on computational assumptions such as the Exponential Time Hypothesis (ETH) which asserts that 3-SAT requires exponential time to solve. Our unconditional results achieve the same parameters as all of these previous results (for separable states) or as some of the previous results (for quantum correlations). In some cases we can make use of the framework of Lee–Raghavendra–Steurer (LRS) to establish integrality gaps for any SDP extended formulation, not only the SoS hierarchy. Our hardness result on separable states also yields a dimension lower bound of approximate disentanglers, answering a question of Watrous and Aaronson et al.

These results can be viewed as limitations on the monogamy principle, the PPT test, the ability of Tsirelson-type bounds to restrict quantum correlations, as well as the SDP hierarchies of Doherty–Parrilo–Spedalieri, Navascues–Pironio–Acin, and Berta–Fawzi–Scholz. Indeed a wide range of past work in quantum information can be described as using an SDP on one of the above two problems and our results put broad limits on these lines of argument.

Contents

1. Introduction . . . 424
1.1 2-to-4 norms and separability of quantum states . . . 424
1.2 Entangled games . . . 429
1.3 Technical contributions . . . 431
1.4 Open problems . . . 433
2. Preliminaries . . . 434
2.1 Polynomial optimization and sum-of-squares proofs . . . 434
2.2 Quantum information . . . 436
2.3 SoS hierarchies for quantum problems . . . 437
3. Framework of Deriving Lower Bounds . . . 438
3.1 Optimization problems and integrality gaps . . . 438
3.2 Reduction between optimization problems . . . 441
3.3 3XOR with integrality gap . . . 444
4. Lower Bounds on h_{Sep} and Its Applications . . . 445
4.1 From 3XOR to 2- OUT- OF- 4- SAT- EQ . . . 445
4.2 From 2- OUT- OF- 4- SAT- EQ to QMA(2)- HONEST . . . 448
4.3 DPS lower bounds from other protocols . . . 452
5. SDP extended formulation Lower Bounds for h_{Sep} . . . 454
5.1 SDP extended formulation lower bounds from LRS . . . 454
5.2 Applying LRS to h_{Sep} . . . 456
5.3 The no-disentangler conjecture . . . 458
6. Lower Bounds for Entangled Games . . . 460
6.1 General SDPs . . . 462
6.2 An explicit lower bound for ncSoS . . . 463
References . . . 465

1. Introduction

1.1. 2-to-4 norms and separability of quantum states.

Separable (unentangled) quantum states Entanglement is an essential ingredient in many applications of quantum information processing. Understanding the characterization of entangled quantum states remains a fundamental problem in quantum information processing research. For example, studying the boundary between entangled and separable quantum states has been useful for a variety of problems in quantum information such as data hiding [DPD02], teleportation [Mas06], privacy [BCHW15], channel capacities [Rai01,MSW04,MW12], and the quantum marginal problem [CJYZ15].

Distinguishing between entangled and separable quantum states, also known as the separability problem, turns out to be closely related to the following optimization problem $h_{\mathrm{Sep}(d,d)}$, defined for a positive-semidefinite $d^2 \times d^2$ matrix M as

$$h_{\mathrm{Sep}(d,d)}(M) := \max_{\substack{x,y\in\mathbb{C}^d \\ \|x\|_2=\|y\|_2=1}} \sum_{i,j,k,l\in[d]} M_{ij,kl} x_i^* x_j y_k^* y_l. \tag{1.1}$$

In general these problems cannot be approximated in polynomial time even to constant error [BBH+12,HM13], assuming the ETH. Let $\mathrm{Sep}^k(d)$ denote the convex hull of $|\psi_1\rangle\langle\psi_1|\otimes\cdots\otimes|\psi_k\rangle\langle\psi_k|$ as $|\psi_1\rangle, \ldots, |\psi_k\rangle$ range over all unit vectors in $\mathbb{C}^d$. ($h_{\mathrm{Sep}(d,d)}$

Table 1. Hardness results for $h_{\mathrm{Sep}^k(d)}$

Reference	k	c	s	n	Notes
[LNN12]	2	1	$1 - \frac{1}{d \cdot \mathrm{poly}\log(d)}$	$O(d)$	(1)
[Per12]	2	1	$1 - \frac{1}{\mathrm{poly}(d)}$	$O(d)$	(2)
[ABD+09]	$\sqrt{d} \cdot \mathrm{poly}\log(d)$	1	0.99	$O(d)$	(3)
[CD10]	$\sqrt{d} \cdot \mathrm{poly}\log(d)$	$1 - 2^{-d}$	0.99	$O(d)$	(4)
[HM13]	2	1	0.01	$\frac{\log^2(d)}{\mathrm{poly}\log\log(d)}$	(5)

(1) This builds on work in [Gur03,BT09,Bei10] which achieved the same result with $s = 1 - 1/\mathrm{poly}(d)$. Related results were found for testing membership in $\mathrm{Sep}^2(d)$ in [Gur03,Liu07,Gha10]. (2) The measurement M can be implemented using a uniform quantum circuit of size $\mathrm{poly}\log(d)$. (3,4,5) Here 0.99 refers to a constant strictly less than 1 whose explicit value is not known, and 0.01 means the result applies for any constant in the range (0, 1). (4) The measurement M can be taken to be a Bell measurement, meaning that all the systems are measured locally and then the answers are processed classically. (5) M can be taken to be separable, i.e. of the form $M = \sum_i A_i \otimes B_i$ for $A_i, B_i \geq 0$

is $h_{\mathrm{Sep}^2(d)}$.) For a general convex set S, let $h_S(x) := \max_{y \in S} \langle x, y \rangle$. Consider the problem of determining whether $h_{\mathrm{Sep}^k(d)}(M)$ is $\geq c$ or $\leq s$ for some $0 \leq s < c \leq 1$ and some matrix M such that $0 \leq M \leq I$. Several hardness results are known of the form "determining satisfiability of 3-SAT instances with n variables and $O(n)$ clauses can be reduced to estimating $h_{\mathrm{Sep}^k(d)}$ in this way." We summarize such hardness results in Table 1.

These results can be thought of as ETH-based no-go results, since in each case ETH implies a lower bound on the run-time of any algorithm approximating h_{Sep}, and in particular implies the existence of integrality gaps for the SoS hierarchy. We mention also one hardness result that does not fit into this framework is the result by [BBH+12] that a constant-factor multiplicative approximation to $h_{\mathrm{Sep}^2(d)}$ could be used to solve Unique Games instances of size $d^{\Omega(1)}$.

DPS hierarchy for separability problem and integrality gaps Despite the worst-case hardness, a variety of heuristics have been developed for the separability problem given the utility of solutions even for specific cases.

The set of entangled states was first approximated by the set of states with non-positive partial transpose [Per96,HHH96]. The resulting test is known as the "PPT test" and it is known that all separable states have positive semidefinite partial transpose (i.e. are PPT) and that some entangled states are PPT while others are not.

Doherty, Parrilo, and Spedalieri improved this to a hierarchy of approximations [DPS04]. The kth level of the so-called DPS hierarchy approximates the set of entangled states by the set of states ρ^{AB} for which there does not exist $\tilde{\rho}^{A_1 \ldots A_k B_1 \ldots B_k}$ with $\rho^{AB} = \tilde{\rho}^{A_1 B_1}$, the supports of $\tilde{\rho}^{A_1 \ldots A_k}$ and $\tilde{\rho}^{B_1 \ldots B_k}$ contained in the symmetric subspace and $\tilde{\rho}$ remaining positive semidefinite under the partial transpose of any set of subsystems. Again all separable states pass the level-k DPS test as do some entangled states. As $k \to \infty$ the run-time increases exponentially but the accuracy also increases, meaning that fewer entangled states pass the test. Since the DPS sets include all separable states as well as some entangled states, we call the approximation a "relaxation" of the set of separable states, and call entangled states which pass the DPS test "integrality gaps".

The accuracy of the DPS hierarchy has been analyzed by a long sequence of works which have found various positive results, matching the barriers from ETH in a few cases [DF80,CFS02,KR05,KM09,CKMR07,BCY11]. A handful of negative results are also

known but generally only for weaker versions of the DPS hierarchy. If we require only that $\tilde{\rho}$ be symmetric (i.e. commute with permutations), then the antisymmetric state is a potent integrality gap showing that this weaker hierarchy still makes large errors until $k > d$. This can also be turned into a integrality gap for the slightly stronger hierarchy which restricts the support of $\tilde{\rho}$ to be contained in the symmetric subspace, but it is easily detected with the PPT test. Another integrality gap is known to defeat the larger class of tests for which $\tilde{\rho}$ is required to be symmetric and PPT across any cut, but only works up to $k = O(\log d)$ [BCY11]. Two recent results regarding the constant-error case are of particular interest are those of Lancien [Lan15] and Aubrun and Szarek [AS15]. The result of Lancien uses a probabilistic argument to show that most d-dimensional non-separable states will pass the k-extendibility test together with the requirement that the state be PPT across a single cut up to $k = O(d)$. This is a better lower bound on the level of relaxation than our result, but only for a weaker version of the DPS hierarchy. The result of Aubrun and Szarek studies the set of positive but not completely positive maps; it is known that any non-separable state can be detected by such a map. Using tools from convex geometry, they show a lower bound of $\exp(\Omega(d^3/\log d))$ for the number of positive maps needed to detect all d-dimensional states that are at least a constant distance away from separable. This can be viewed as a lower bound on the dimension of a specific class of SDPs, which have a constraint of the form $\mathrm{Id} \otimes (\Phi_1(\rho) \oplus \Phi_2(\rho) \oplus \cdots \oplus \Phi_N(\rho)) \succeq 0$, where $\Phi_1 \ldots \Phi_N$ are positive maps on d-dimensional matrices. This class of SDPs contains the PPT test as well as many other SDPs, but it does not contain higher levels of DPS.

All of these integrality gaps except that of [Lan15] are known to be defeated by the full DPS hierarchy and there is no known way to modify them to avoid this. Indeed the only previously known unconditional negative result was in the original DPS paper which showed that the error always remained nonzero for all finite values of k (see also [BS10] showing that this could be amplified). Indeed one can even define an improved version of DPS that removes this limitation and always exactly converges at a finite (but large) value of k [HNW15].

We remark that the goal of much of this past work has been much more general than separability testing. The convergence proofs of DPS are related to "monogamy relations" which bound how widely an entangled state can be shared. These in turn have been related to the security of quantum key distribution, rigorous proofs of the mean-field approximation in many-body physics [BH13a], quantum interactive proofs and many other applications. In some cases also integrality gaps for weaker hierarchies have been useful examples of extremal information-theoretic behavior [CSW12].

Semidefinite programming (SDP) and sum-of-squares (SoS) hierarchies The DPS hierarchy is a special case of a more general approach to polynomial optimization problems known as the Sum-of-Squares (SoS) hierarchy. The SoS hierarchy in turn is an example of a Semidefinite Programming (SDP) relaxation [BV04]. A major question in the theory of algorithms and complexity is the power of SDP relaxations and in particular the SoS hierarchy [BS16]. Most problems in NP admit various SDP relaxations, but the worst-case quality of the resulting approximations is often unknown.

The SoS hierarchy was introduced in [Sho87,Nes00,Par00,Las01] and reviewed in [Lau09,Bar14]. It is a family of SDP relaxations, parametrized by the problem size n and the level of the hierarchy k. They run in time $n^{O(k)}$ and generally converge to the correct answer as $k \to \infty$, but a crucial question is to determine this rate of convergence. Over several domains, we know that when k is $\Omega(n)$, convergence is achieved: at $k = n$ exactly

when the domain is the boolean hypercube, and approximately when the domain is the hypersphere [DW12,BGL17]. In this regime, the runtime of the algorithm is comparable to that of brute-force search. However, in some cases, stronger positive results are known for $k = O(1)$ or $k = O(\log n)$.

On the lower bound side, integrality gaps are known for which the SoS hierarchy, or in some cases, more general families of SDPs, fail to give the correct answer. Integrality gaps are known even for problems that are easy to solve, such as linear equations over a finite field [Gri01]. Our first family of results can be described as a set of integrality gaps for the DPS hierarchy.

Unique games, small-set expansion and the 2-to-4 norm One further application of our results is to problems that are not obviously related to quantum mechanics. A central question in the theory of approximation algorithms is the unique games conjecture (UGC, introduced in [Kho02] and reviewed in [Tre12]) which asserts the NP-hardness of a problem known as the "unique games problem." If true, the UGC would imply that level $k = 1$ of the SoS hierarchy achieves optimal approximation ratios for a wide range of problems. There is a subexponential-time algorithm for the unique games problem, and no $k = \omega(1)$ SoS integrality gap instances are known. While these cast doubt on the UGC, there is other evidence in favor, including a conjectured $n^{\Omega(1)}$-round integrality gap and the fact that slight modifications of the approximation parameters are known to yield NP-hard problems.

A close variant of the UGC is the small-set expansion (SSE) hypothesis which asserts that $\forall \eta > 0, \exists \delta > 0$ such that it is NP-hard to determine whether the maximum expansion of all subsets of fractional size δ is $\leq \eta$ or $\geq \eta$. The SSE hypothesis implies UGC and indeed SSE is equivalent to a slightly restricted form of the unique games problem [RST12]. Here too no $k = \omega(1)$ integrality gaps are known.

The SSE problem in turn can be relaxed to the problem of estimating the $2 \to 4$ norm of a matrix. If $A \in \mathbb{R}^{m \times n}$ then define the $2 \to 4$ norm of A to be

$$\|A\|_{2\to 4} := \max_{x \neq 0} \frac{\|Ax\|_4}{\|x\|_2} \qquad \text{where} \qquad \|y\|_p := \left(\sum_i |y_i|^p \right)^{1/p}.$$

(Here the number "4" can be replaced by any constant $q > 2$.)

The $2 \to 4$ norm of a matrix is closely related to h_{Sep}. Indeed if we set $M_{ij,kl} := \sum_a A_{ia} A_{ja} A_{ka} A_{la}$ for a real matrix A then a straightforward calculation shows that $h_{\mathrm{Sep}(d,d)}(M) = \|A\|_{2\to 4}^4$, implying that computing $\|A\|_{2\to 4}$ reduces to computing $h_{\mathrm{Sep}}(\cdot)$. Conversely h_{Sep} can be reduced to calculating a $2 \to 4$ norm but this requires somewhat more work [BBH+12].

It was also shown in [BBH+12] that SSE for a graph G is approximately related to the $2 \to 4$ norm of the projector onto the top eigenspace of G. Thus, algorithms for the $2 \to 4$ norm yield algorithms for SSE, and indeed known positive results for the SoS hierarchy translate into subexponential time algorithms for SSE [BBH+12] (matching those found using other methods). On the other hand, the quasipolynomial hardness known for the $2 \to 4$ norm does not necessarily imply similar hardness for SSE. (This latter hardness result assumes the Exponential-Time Hypothesis (ETH) [IP01] which asserts that 3-SAT instance with n variables require $2^{\Omega(n)}$ time to solve.)

Before our work, no $k = \omega(1)$ SoS integrality gap was known for the $2 \to 4$ norm, and finding such an integrality gap was proposed as an open problem by Barak [Bar14]. One of our main results is to establish SoS integrality gaps for the $2 \to 4$ norm problem

which roughly match the known computational hardness results, but without needing the assumption of the ETH. We also establish weaker integrality gaps for a more general class of SDP relaxations, called SDP extended formulations.

Our contributions Our main contribution is to provide instances on which the above hierarchies (and others) fail to give the correct answer. Thus we give lower bounds that do not rely on complexity-theoretic assumptions (and are in that sense "unconditional"), although they do only apply to a specific family of algorithms. Additionally, our SoS bounds provide an explicit integrality gaps for the DPS hierarchy, while the ETH-based hardness result can only imply its existence.

Our results show (unconditionally) that the DPS hierarchy cannot estimate $h_{\mathrm{Sep}^2(d)}$ to constant accuracy at level k unless $k \geq \tilde{\Omega}(\log d)$, corresponding to run-time $d^{\tilde{\Omega}(\log d)}$. Similarly the $2 \to 4$ norm of d-dimensional matrices cannot be approximated to within a constant multiplicative factor using $O(\log(d)/\mathrm{poly}\log\log(d))$ levels of SoS. These results are described in Corollaries 4.8 and 4.9. This yields the first unconditional quantitative limits to the rate at which the DPS hierarchy converges, and likewise the first $k = \omega(1)$ SoS integrality gaps for the $2 \to 4$ norm problem.

We also demonstrate integrality gaps corresponding to the other rows of Table 1; In some cases these are nearly tight. In Theorem 4.10, we show that accuracy $\tilde{O}(1/d)$ [LNN12,NOP09] requires $\tilde{\Omega}(d)$ levels of the SoS hierarchy, corresponding to exponential time. Likewise for $h_{\mathrm{Sep}^{\sqrt{n}}(n)}$, constant-error approximations are shown in Theorem 4.11 to require $\Omega(n)$ levels, which matches known SoS achievability results [BH13b].

We further apply techniques due to Lee, Raghavendra, and Steurer [LRS15] to extend our no-go theorems to a more general class of relaxations called SDP extended formulations. This is the first extension of LRS that we are aware of to non-boolean domains. To do this we need to work with functions that are not strictly self-reducible, which requires some modifications of the techniques of LRS. However, while these bounds cover a larger class of SDPs than the above bounds, they are quantitatively weaker.

Theorem 1.1. (Informal, refer to Theorem 5.6 and Sect. 5) *Any SDP extended formulation of* $h_{\mathrm{Sep}^2(d)}$ *achieving accuracy* $1/\mathrm{poly}(d)$ *must have a total number of variables* $\geq d^{\tilde{\Omega}(\log d)}$.

Here "SDP extended formulation" is a technical condition defined in the body of our paper. Roughly speaking an SDP extended formulation should replace the optimization over separable states with an optimization over a convex superset, which is defined by a semidefinite program. (In fact we rule out a slightly larger class of approximations.)

One corollary of these results is an unconditional proof of a version of the "no approximate disentangler" conjecture of Watrous,[1] for which previously only the zero-error case was known [ABD+08]. This conjecture asserts that if $\mathcal{N}$ is a quantum channel from D dimensions to $d \times d$ dimensions such that $\mathrm{Sep}(d, d) \approx \mathrm{Image}(\mathcal{N})$ (with "$\approx$" defined precisely in the body of our paper, but roughly speaking it corresponds to $1/\mathrm{poly}(d)$ error) we must have $D \geq d^{\tilde{\omega}(\log(d))}$.[2] Our results imply unconditional lower bounds on

[1] The original goal of the conjecture was to rule out a particular strategy for putting QMA(2) inside QMA. Here QMA(2) is the set of languages where membership can be verified using a proof that is a pair of unentangled quantum states.

[2] This is not the strongest possible version of the conjecture since one could conceivably demand that $D \geq \exp(d)$; this stronger form is false for the 1-LOCC norm [BCY11] but is still an open question for trace distance.

Table 2. Hardness results for $\omega_{\text{entangled}}$ with k provers and question alphabet size Q

Reference	k	c	s	n
[KKM+11]	3	1	$1 - \frac{1}{\text{poly}(Q)}$	$O(Q)$
[IKM09]	2	1	$1 - \frac{1}{\text{poly}(Q)}$	$O(Q)$
[IV12]	4	1	$2^{-Q^{\Omega(1)}}$	$Q^{\Omega(1)}$
[Vid13]	3	1	$2^{-Q^{\Omega(1)}}$	$Q^{\Omega(1)}$

the input dimension D albeit with parameters somewhat weaker than those based on ETH.

Theorem 1.2 (Informal, refer to Theorem 5.9). *Let* $D = \dim(\mathcal{H}), d = \dim(\mathcal{K})$, *and suppose that* $\Lambda : D(\mathcal{H}) \to D(\mathcal{K} \otimes \mathcal{K})$ *is an approximate disentangler with* $1/\text{poly}(d)$ *error. Then*

$$D \geq d^{\tilde{\Omega}(\log(d))}.$$

1.2. Entangled games.

Noncommutative polynomial optimization and entangled games Another major class of optimization problems concerns polynomials in non-commuting variables [HP06, PNA10]. As we explain in Sect. 2.1, these involve optimizing operator-valued variables over a vector space of unbounded or even infinite dimension. One application is to understanding the set of "quantum correlations", meaning the conditional probability distributions $p(x, y|a, b)$ achievable by local measurements on a shared entangled states [NPA08]. The most famous example of a quantum-but-not-classical correlation was discovered by Bell in 1964 [Bel64] and gave a concrete experiment for which quantum mechanics predicts outcomes that are incompatible with any theory that lacks entanglement or faster-than-light signaling. More recently, quantum correlations are studied in the context of multi-prover games where the provers share entanglement (i.e., *entangled games*); here, without a bound on the dimension of the shared entangled state, one cannot rule out even infinite-dimensional systems. Non-commuting optimization can be useful even in cases where the dimension is finite but exponentially large and the goal is to obtain smaller optimization problems, e.g. in quantum chemistry [Maz04].

A similar hierarchy was developed to approximate the set of correlations achievable with local measurements of quantum states [NPA08, DLTW08, BFS15], or for more general polynomial optimization [PNA10]. This is known variously as the noncommutative Sum of Squares (ncSoS) hierarchy or the NPA (Navascues-Pironio-Acin) hierarchy. Here, much less is known on either the positive or negative side. The ncSoS hierarchy similarly has complexity increasing exponentially with k and similarly converges as $k \to \infty$, although it is a famous open question (Tsirelson's problem [SW08]) whether it indeed converges to the value of the best quantum strategy.

Computational hardness results are also known for the entangled value of quantum games. If $\omega_{\text{entangled}}$ refers to the entangled value of a game with k provers, one round, questions in $[Q]$, answers in a $O(1)$-sized alphabet, completeness c and soundness s, then the known reductions from 3-SAT instances of size n are described in Table 2. However, no unconditional results are known for $k > 5$.

Despite the lack of general results, specific solutions to the ncSoS hierarchy can be extremely useful. For example, Tsirelson's 1980 outer bound on the winning probability of a particular quantum game [Cir80,CB96] has since had widespread application to topics including communication complexity [BBL+06], "self-testing" quantum systems and multiparty secure quantum computing [BP15,RUV13a,RUV13b], and device-independent cryptography [BHK05]. In quantum chemistry, the Pauli exclusion principle can be expressed as an operator inequality [CM15], and far-reaching generalizations exist [AK08], each of which can be seen as a dual feasible point for a ncSoS hierarchy. While [AK08] used representation theory to show the existence of these points in general, even explicit specific solutions can be useful [Rus07,Kly13].

Our contributions We show the first known limitations on the ncSoS hierarchy for $k = \omega(1)$.

Theorem 1.3 (Informal, refer to Theorem 6.11). *There exists a sequence of two-player non-local games G_n such that the entangled game value $\omega_{entangled}(G_n) \leq 1 - c/n^2$ for some constant c but the ncSOS hierarchy believes $\omega_{entangled}(G_n) = 1$ up to level $m = \Omega(n)$.*

Previously the only unconditional lower bound on the error of the ncSoS hierarchy for large constant k was Slofstra's uncomputability result [Slo16], which implied (among other things) that the $\epsilon = 0$ case could not be solved by the kth level of the ncSoS hierarchy unless k is an uncomputably large function of the inputs.
We are also able to give dimension lower bounds for any SDP extended formulation achieving reasonable accuracy, although these do not match our SoS lower bounds.

Theorem 1.4 (Informal, refer to Theorem 6.9). *There exists a sequence of two-player non-local games G_n such that any SDP extended formulation approximating the entangled game value $\omega_{entangled}(G_n)$ to precision $O(1/n^2)$ has dimension $\geq n^{\log n/\text{poly}\log\log n}$.*

As in the case of h_{Sep}, we obtain SDP lower bounds by using the technique of [LRS15]. It is interesting to note that the games setting is somewhat more amenable to LRS's techniques than h_{Sep}. This is because a key step in the proof of LRS requires embedding hard instances of the target problem into instances of the same problem with many more variables (this enables the use of random restrictions, which is crucial to the LRS proof). In the case of games, it is easy to embed a game with a smaller question alphabet into a game with a larger one—the referee simply ignores the extra questions. In contrast, this is not so simple for h_{Sep}, since most QMA(2) protocols for CSPs involve sampling from a state which is a uniform superposition over all the variables, and most samples will not lie within a given small subset of the variables.

Our results here do not fully match the known computational hardness results. In particular, the ETH-based arguments work at constant accuracy and our results rule out only approximations whose error decreases as a power of the number of questions and answers. This is because the reductions that yield constant accuracy ETH-conditional hardness involve operations like low-degree polynomial testing over higher order finite fields, which are not easy to make SoS complete. At the same time, it is worth noting that Theorem 6.9 extends to other SDPs for quantum correlations and in particular also limits the stronger hierarchy described in [BFS15].

1.3. Technical contributions.

Obtaining SoS hardness results At a high level, similar to many other SoS hardness results, all of our SoS hardness results stem from a classic result of Grigoriev [Gri01], showing hardness for the problem 3XOR in the SoS model. To obtain integrality gaps for the problems considered here, we need to reduce (and sometimes embed) the classical hard problems in the SoS model to quantum problems, because our derivation of the SoS hardness of the 2-to-4 norm is inspired by its connection to the separability problem in quantum information.

There have been several previous examples using reductions to prove the hardness in the SoS model (e.g., [Tul09,OWWZ14]). In Sect. 3, we formulate a framework of "low-degree reductions" which can be used in many cases. It helps us to revisit previous NP-hardness results and prove that they extend to yield integrality gaps.

We hope that this framework could facilitate the proof of hardness in both the SoS model and the more general SDP model. Another feature of this framework is that it can easily go beyond the problems over boolean domains. As far as we know, our later application of such a framework derives the first SoS hardness results for problems over *non-boolean* domains.

Let us look more closely at how this framework works. To obtain integrality gaps in the SoS model, one needs to show that (a) the SoS solution believes the value is large up to high level (degree), and (b) the true value is actually small. To achieve (a), we introduce the notion of *low-degree reductions*, in which one requires the reductions preserve a SoS solution for the reduced problems with almost the same value and a small amount of loss of the degree. These goals of preserving SoS solutions were referred to as "Vector Completeness" in [Tul09] and "SoS Completeness" in [OWWZ14]. In those papers this property was shown using direct analysis of the SDP solutions. Our approach is instead to show that along with the NP-hardness reductions there exist low-degree polynomials mapping solutions of the original problem to solutions of the new problem. The more complicated map on SDP solutions then follows using generic arguments about low-degree reductions.

To achieve (b), we resort to the study of quantum interactive proof protocols, which are connected to the separability problem, hence 2-to-4 norms, and at the same time provide the NP-hardness reductions shown in Table 1. In particular, we can import results such as the QMA(2) protocol for 3-SAT problem by Aaronson et al. [ABD+08] and Chen-Drucker [CD10], and make use of the *soundness* of these reductions to achieve (b). We follow the similar principle to import the two-player non-local game protocol from Ref. [IKM09] (also in Table 2) to achieve (b) for SoS hardness results about entangled games.

There is an issue with directly adopting Aaronson et al.'s QMA(2) protocol for 3-SAT as the reduction, because the first step of this reduction makes use of the PCP theorem, which turns out to be high-degree. The use of the PCP theorem is to amplify the gap between promised instances with some additional features. The SoS pseudo-solution for 3XOR has already satisfied part of these features. We replace this step by a direct and low-degree construction inspired by the proof of the PCP theorem.

As a minor technical contribution, we observe and make great use of the correspondence between density operators in quantum information and pseudo-expectations in SoS (e.g., Definition 4.6). This correspondence helps simplify a lot of calculations, especially given that our reductions are protocols from the study of quantum interactive proofs. We hope this connection can be found useful beyond the scope of this paper.

Extending to SDP extended formulations Extending the above SoS hardness to the more general class of SDP extended formulations crucially relies on the recent breakthrough result by by Lee, Raghavendra and Steurer [LRS15]. Despite being by far the most successful route to obtain unconditional bounds on general family of SDP relaxations, the LRS result has a few serious limitations that make it difficult to extend general SoS hardness result (see Sect. 5.1 for technical details.)

1. The result of LRS only works on the boolean cube $\{0, 1\}^n$ and applies to CSP problems. Recent works by Braun et al [BPZ15] demonstrate the possibility of extending LRS's result to non-CSP problems (such as VertexCover, Max-Multi-CUT) but still over boolean domains, through affine reductions (or some relaxation of affine reductions discussed in [BPR16]).
   It is, however, not a priori clear whether this approach can be extended to non-boolean domains. Even for finite domains, the proof of [LRS15] may not automatically generalize as it makes intricate use of the structure of $\{0, 1\}^n$, in particular Fourier analysis of functions over this domain. It is rather less clear how to extend the LRS approach to the continuous domains considered here—the hypersphere, for the separability problem, and infinite-dimensional Hilbert spaces, for the entangled game problem—without making use of, e.g., the theory of spherical harmonics.
   On top of that, most reductions used in deriving our SoS hardness are low-degree rather than affine (as in [BPZ15,BPR16]). Finally, the particular class of relaxations that [LRS15] applies to, here referred to as SDP extended formulations, must satisfy certain rather rigid technical constraints, in particular one which we formulate as the embedding property.
   Our crucial observation is that our reductions inspired by quantum protocols actually allow one to embed hard problems over $\{0, 1\}^n$ into larger, even infinite-dimensional, domains. Moreover, our reductions are naturally low-degree, which could be a technically interesting point comparing to the affine reductions in [BPZ15,BPR16]. As a result, we can avoid extending [LRS15]'s analysis to each setting, while still extending its results on $\{0, 1\}^n$ to much larger domains, albeit with some loss in parameters.
2. The SDP lower bounds obtained by this result will never be better than quasi-polynomial in the problem size because of technical constraints. This limitation seems essential for the random restriction analysis that is central to the proof of LRS. (Recent work [KMR17] has succeeded in removing this limitation in the special case of LP relaxations, but doing so for the full class of SDP relaxations considered by LRS remains open.) A unfortunate consequence is that the lower bounds on general SDPs obtained via this method can be much looser than the SoS lower bounds they are based on. This partially explains why we have stronger SoS lower bounds than general SDP lower bounds in our results.
3. The LRS lower bound also crucially relies on a certain "self-reducible" structure of the problems. This is because of the use of pattern matrices in the proof of LRS. Intuitively, LRS can only show that some problem of size n is hard if one can simultaneously embed a hard instance of size $m < n$ into each size-m subset of the size-n input. This is the case for CSP problems as well as for the problems considered in [BPZ15,BPR16].
   However, because of our use of quantum protocols that involve superposed quantum states over all of the input variables, it is no longer true that a similar requirement will be satisfied for the quantum problems. As a result, we will only be able to obtain SDP lower bounds in *some* of the cases where we have SoS lower bounds, and even in those cases, our parameters will be worse than those of the SoS results.

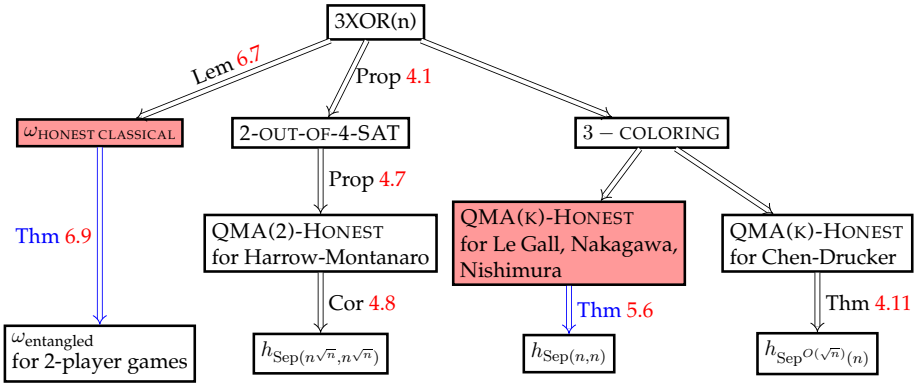

Fig. 1. All our results are derived by applying a series of low-degree reductions to the integrality gap for 3XOR given by [Gri01]. The red nodes indicate problems over the boolean cube to which the LRS theorem is applied. The blue arrows are "embedding reductions" that map problems over the boolean cube to problems over general domains, e.g. the set of separable states or the set of quantum entangled strategies

In summary, to the authors' knowledge, we are the first to extend LRS to problems over non-boolean domains. However, due to the above limitations of LRS, there are gaps between what we could get for SoS lower bounds and for SDP lower bounds. We consider it a major open problem to improve on these techniques and prove tighter SDP lower bounds.

Handling SoS with non-commutative variables As our last technical contribution, we also demonstrate how to derive an ncSoS solution from a SoS solution for the purpose of showing the hardness of approximating entangled non-local game values. This step is necessary as all our SoS hardness results stem from the 3XOR problem, even for the optimization problem over non-commutative variables in the context of entangled games. As a result, without the following step, we only obtain a SoS hardness result rather than an ncSoS hardness result for entangled games.

Our idea is to embed a SoS solution into an ncSoS solution, where the embedding is due to the connection between each question and a prover's corresponding strategy (i.e., operators). Intuitively the ncSoS solution constructed in this way can be said to "cheat" in the same way that classical SoS solutions do, and not by exploiting entanglement the way a "valid" quantum strategy would. We refer curious readers to the proof of Theorem 6.11 for details.

We summarize the complete reductions that handle all technical details in Figure 1.

1.4. Open problems. Our work leaves open a number of intriguing questions.

- Our results can be viewed as an extension of [LRS15] beyond $\{0, 1\}^n$ to sets such as the hypersphere. However, as explained in the above, we face a lot of difficulties in doing so because of the limitations of LRS. We consider it a major open problem to improve on these techniques and prove tighter SDP lower bounds over general domains. In particular, we believe that the quantum problems considered in this paper are a good motivation to prove extension complexity lower bounds for CSPs that are exponential, and not just superpolynomial. Progress in this direction has been achieved for LP extension complexity [KMR17], but the SDP case remains open.

- It raises the motivation to examine convex but non-SDP-based relaxations, such as the entropic bounds used in [PH11].
- Our integrality gaps for the noncommutative hierarchies involve games whose entangled values cannot be effectively bounded using low-degree SoS proofs, but can be bounded using other methods. These methods usually rely on specific features of the game; e.g. consistency checks or other tests [IKM09,IV12]. Can these upper-bound methods be understood more generally? For example, can they be described using a high level of the SoS hierarchy? A motivating example is 3XOR, where Gaussian elimination can be used to refute unsatisfiable instances once we get to level n of the hierarchy [Gri01]. We note that, contrary to Barak's "Marley principle" [Bar14], the soundness bounds in [IKM09,IV12] neither rely on the probabilistic method nor low-degree polynomials.
- Our results on quantum correlations should be strengthened to match the ETH-based bounds (e.g. [IV12]). We also give new motivation for proving that random 3XOR has low entangled value (cf. section 2.3 of [Pal15]). If known, this would give a much more direct and efficient no-go result for games.
- We have constructed entangled states that appear separable to the lower levels of the DPS hierarchy. But are these states generic in the sense of [Lan15]? We believe that this is the case; i.e. a random state within the convex set accepted by DPS is likely to have distance from separable states that is comparable or better to our examples. However, different techniques will be needed to prove this.

2. Preliminaries

We provide a brief introduction to sum-of-squares proofs/optimization in Sect. 2.1 and then summarize relevant background about quantum information and our terminology in Sect. 2.2.

2.1. Polynomial optimization and sum-of-squares proofs. In this section, we lay out the basics of the sum-of-squares (SoS) optimization algorithms. They were introduced in [Sho87,Nes00,Par00,Las01] and reviewed in [Lau09,Bar14].

Symmetric subspace. For any vector space V and positive integer k, the permutation group on k letters S_k acts on the vector space $V^{\otimes k}$ by $P_\pi \left|i_1\right\rangle \otimes \ldots \otimes \left|i_k\right\rangle = \left|i_{\pi^{-1}(1)}\right\rangle \otimes \ldots \otimes \left|i_{\pi^{-1}(k)}\right\rangle$. The k-partite symmetric subspace $\mathrm{Sym}^k V$ is the subspace of $V^{\otimes k}$ that is invariant under this action:

$$\mathrm{Sym}^k V = \{\left|v\right\rangle \in V^{\otimes k} : P_\pi \left|v\right\rangle = \left|v\right\rangle \ \forall \pi \in S_k\}.$$

Polynomials and non-commutative polynomials. Let $\mathbb{R}[x] := \mathbb{R}[x_1, \ldots, x_n]$ be the set of real-valued polynomials over n variables, and let $\mathbb{R}[x]_d$ be the subspace of polynomials of degree $\leq d$. We also define $\mathbb{R}\langle X\rangle$ to be the set of *non-commutative polynomials* in $X_1, \ldots, X_n$, which we think of as Hermitian operators that do not necessarily commute, and which we call "nc polynomials" for short. The nc polynomials of degree $\leq d$ are denoted $\mathbb{R}\langle X\rangle_d$ and are isomorphic to $\bigoplus_{d' \leq d} (\mathbb{R}^n)^{\otimes d'}$, while the ordinary commutative polynomials $\mathbb{R}[x]_d$ can be viewed as $\bigoplus_{d' \leq d'} \mathrm{Sym}^{d'} \mathbb{R}^n$, where $\mathrm{Sym}^{d'} V$ denotes the symmetric subspace of $V^{\otimes d'}$. These notational conventions follow the optimization literature with the roles of n and d reversed from the way they are usually used in quantum information theory.

Polynomial optimization. Given polynomials $f, g_1, \ldots, g_m \in \mathbb{R}[x]$, the basic polynomial optimization problem is to find

$$f_{\max} := \sup_{x \in \mathbb{R}^n} f(x) \text{ subject to } g_1(x) = \cdots = g_m(x) = 0. \tag{2.1}$$

Equivalently we could impose inequality constraints of the form $g_i'(x) \geq 0$ but we will not explore this option here. In the non-commutative setting we need to optimize over both variables and a density. Given $F, G_1, \ldots, G_m \in \mathbb{R}\langle X \rangle$, define

$$F_{\max} := \sup_{\rho, X=(X_1,\ldots,X_n)} \operatorname{tr}[\rho F(X)] \text{ subject to } \rho \geq 0, \operatorname{tr} \rho = 1, G_1(X) = \cdots = G_m(X) = 0. \tag{2.2}$$

Note that the supremum here is over density operators ρ and Hermitian operators $X_1, \ldots, X_n$ that may be infinite dimensional; see [SW08] for a discussion of some of the mathematical difficulties here.

Sum-of-Squares (SoS) proofs. Although (2.1) and (2.2) are in general NP-hard to compute exactly, the SoS hierarchy is a general method for approximating $f_{\max}$ or $F_{\max}$ from above. This complements simply guessing values of x or (ρ, X) which provides lower bounds on $f_{\max}$ or $F_{\max}$ when they satisfy the constraints. Focusing for now on the commuting case, a SoS proof is a bound that makes use of the fact that $p(x)^2 \geq 0$ for any $p \in \mathbb{R}[x]$. In particular, a SoS proof that $f(x) \leq c$ for all valid f is a collection of polynomials $p_1, \ldots, p_k, q_1, \ldots, q_m \in \mathbb{R}[x]$ such that

$$c - f = \sum_{i=1}^{k} p_i^2 + \sum_{i=1}^{m} q_i g_i. \tag{2.3}$$

Observe that the RHS is ≥ 0 when evaluated on any x satisfying $g_i(x) = 0, \ \forall i$; for this reason, we refer to (2.3) as a Sum-of-Squares (SoS) proof. In particular, it is a proof that $c - f(x) \geq 0$ whenever $g_i(x) = 0$ for all i. This is a degree-d SoS proof if each term p_i^2 and $q_i g_i$ is in $\mathbb{R}[x]_d$. Finding an SoS proof of degree $\leq d$ can be done in time $n^{O(d)} m^{O(1)}$ using semidefinite programming [Lau09].

If we find the minimum c for which (2.3) holds, then we obtain a hierarchy of upper bounds on $f_{\max}$, referred to as the SoS hierarchy or the Lasserre hierarchy. Denote this upper bound by f_{SoS}^d. Given mild assumptions on the constraints $g_1, \ldots, g_m$ one can prove that $\lim_{d\to\infty} f_{\text{SoS}}^d = f_{\max}$ [Lau09]. The tradeoff between degree d and error $(f_{\text{SoS}}^d - f_{\max})$ is the key question about the SoS hierarchy. We can also express this tradeoff by defining $\deg_{\text{SOS}}(c - f)$ to be the minimum d for which we can find a solution to (2.3). Note that $\deg_{\text{SOS}}$ has an implicit dependence on the $g_1, \ldots, g_m$.

A non-commutative SoS proof can be expressed similarly as

$$c - F = \sum_{i=1}^{k} P_i^\dagger P_i + \sum_{i=1}^{m} Q_i G_i R_i, \tag{2.4}$$

for $\{P_i\}, \{Q_i\}, \{R_i\} \subset \mathbb{R}\langle X \rangle$. Likewise the best degree-d ncSoS (noncommutative SoS) proof can be found in time $n^{O(d)} m^{O(1)}$, and we denote the corresponding value by F_{SoS}^d. It is known that $F_{\max} \leq F_{\text{SoS}}^d$ for all d and $\lim_{d\to\infty} F_{\text{SoS}}^d = F_{\max}$ [HM04].

Pseudo-expectations. We will work primarily with a dual version of SoS proofs that have an appealing probabilistic interpretation. A degree-d pseudo-expectation $\tilde{\mathbb{E}}$ is an element of $\mathbb{R}[x]_d^*$ (i.e. a linear map from $\mathbb{R}[x]_d$ to $\mathbb{R}$) satisfying

- **Normalization**. $\tilde{\mathbb{E}}[1] = 1$.
- **Positivity**. $\tilde{\mathbb{E}}[p^2] \geq 0$ for any $p \in \mathbb{R}[x]_{d/2}$.

We further say that $\tilde{\mathbb{E}}$ satisfies the constraints $g_1, \ldots, g_m$ if $\tilde{\mathbb{E}}[g_i q] = 0$ for all $i \in [n]$ and all $q \in \mathbb{R}[x]_{d-\deg(g_i)}$. Then SDP duality implies that

$$f_{\mathrm{SoS}}^d = \max\{\tilde{\mathbb{E}}[f] : \tilde{\mathbb{E}} \text{ is a degree-}d \text{ pseudo-expectation satisfying } g_1, \ldots, g_m\}. \tag{2.5}$$

The term "pseudo-expectation" comes from the fact that for any distribution μ over R^n we can define a pseudo-expectation $\tilde{\mathbb{E}}[f] := \mathbb{E}_{x\sim\mu}[f(x)]$. Thus the set of pseudo-expectations can be thought of as the low-order moments that could come from a "true" distribution μ or could come from a "fake" distribution. Indeed an alternate approach (which we will use only in Lemma 3.15) proceeds from defining "pseudo-distributions" that violate the nonnegativity condition of probability distributions but in a way that cannot be detected by looking at the expectation of polynomials of degree $\leq d$ [LRS15]. We can define a noncommutative pseudo-expectation $\tilde{\mathbb{E}} \in \mathbb{R}\langle X\rangle_d^*$ similarly by the constraints $\tilde{\mathbb{E}}[1] = 1$ and $\tilde{\mathbb{E}}[p^\dagger p] \geq 0$ for all $p \in \mathbb{R}\langle X\rangle_{d/2}$.

The boolean cube. Throughout this work, we will be interested in the special case of pseudo-expectations over the boolean cube $\{\pm 1\}^n$. This set is defined by the constraints $x_i^2 - 1 = 0$, $i = 1, \ldots, n$, and thus we say that $\tilde{\mathbb{E}}$ is a degree-d pseudo-expectation over $\{\pm 1\}^n$ if for any variable x_i and polynomial q of degree at most $d - 2$,

$$\tilde{\mathbb{E}}[(x_i^2 - 1)q] = 0. \tag{2.6}$$

This means we can define $\tilde{\mathbb{E}}$ entirely in terms of its action on multilinear polynomials.

2.2. *Quantum information.*

Quantum States The state space $\mathcal{A}$ of m-qubit states is the complex Euclidean space $\mathbb{C}^{2^m}$. An m-qubit quantum state is represented by a density operator ρ, i.e., a positive semidefinite matrix with trace 1, over $\mathcal{A}$. The set of all quantum states in $\mathcal{A}$ is denoted by $\mathrm{Dens}(\mathcal{A})$. A quantum state ρ is called a *pure* state if $\mathrm{rank}(\rho) = 1$; otherwise, ρ is called a *mixed* state. An m-qubit pure state $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a unit vector in $\mathbb{C}^{2^m}$. (We might abuse $|\psi\rangle$ for $|\psi\rangle\langle\psi|$ when it is clear from the context.) The Hilbert-Schmidt inner product on the operator space $\mathrm{L}(\mathcal{A})$ is defined by $\langle X, Y\rangle = \mathrm{tr}(X^\dagger Y)$ for all $X, Y \in \mathrm{L}(\mathcal{A})$, where $\dagger$ is the adjoint operator.

An important operation that can be performed on bipartite or multipartite states is the *partial trace*. For a bipartite state $\rho \in \mathrm{Dens}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the partial trace over system A is a density matrix $\rho_B := \mathrm{Tr}_A[\rho]$, whose matrix elements are given by

$$\langle i|\rho_B|j\rangle = \sum_k \langle k| \otimes \langle i|\rho|k\rangle \otimes |j\rangle .$$

One can analogously compute the partial trace of a multipartite state over any subset of its component subsystems. The state obtained by partial tracing some of the subsystems is called the *reduced* state on the remaining subsystems.

Let Σ be a finite nonempty set of *measurement outcomes*. A *positive-operator valued measure (POVM)* on the state space $\mathcal{A}$ with outcomes in Σ is a collection of positive semidefinite operators $\{P_a : a \in \Sigma\}$ such that $\sum_{a\in\Sigma} P_a = \mathsf{Id}_{\mathcal{A}}$. When $P_a^2 = P_a$ for all $a \in \Sigma$, such a POVM is called *projective* measurement. When this POVM is applied to a quantum state ρ, the probability of each outcome $a \in \Sigma$ is $\langle \rho, P_a \rangle$. When outcome a is observed, the quantum state ρ becomes the state $\sqrt{P_a}\rho\sqrt{P_a}/\langle \rho, P_a \rangle$.

Distance Measures For any $X \in \mathrm{L}(\mathcal{A})$ with singular values $\sigma_1, \ldots, \sigma_d$, where $d = \dim(\mathcal{A})$, the trace norm of $\mathcal{A}$ is $\|X\|_1 = \sum_{i=1}^{d} \sigma_i$. The *trace distance* between two quantum states ρ_0 and ρ_1 is defined to be

$$\frac{1}{2}\|\rho_0 - \rho_1\|_1 .$$

Multipartite states and separability Suppose $\mathcal{H}_A$ and $\mathcal{H}_B$ are two state spaces with dimensions d_A, d_B, and consider the bipartite state space $\mathcal{H}_A \otimes \mathcal{H}_B$ obtained by taking their tensor product. It is clear that for any $\rho_A \in \mathrm{Dens}(\mathcal{H}_A)$ and $\rho_B \in \mathrm{Dens}(\mathcal{H}_B)$, $\rho_A \otimes \rho_B$ is a density matrix over $\mathcal{H}_A \otimes \mathcal{H}_B$. However, most density matrices in $\mathrm{Dens}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *cannot* be written in this form as a tensor product. We call any state in the convex hull of the tensor product states *separable*, and all other states *entangled*. More formally, define the set of separable states to be

$$\mathrm{Sep}(d_A, d_B) = \mathrm{conv}(\{\rho \otimes \sigma : \rho \in \mathcal{H}_A, \sigma \in \mathcal{H}_B\}).$$

The problem $h_{\mathrm{Sep}(d_A,d_B)}(M)$ is, given Hermitian M acting over $\mathcal{H}_A \otimes \mathcal{H}_B$ with $\|M\| \leq 1$, compute

$$h_{\mathrm{Sep}(d_A,d_B)}(M) := \max_{\rho\in\mathrm{Sep}(d_A,d_B)} \mathrm{Tr}[M\rho].$$

It is easy to see that the optimum value will be achieved on the extreme points of the separable states, which are simply the pure product states (i.e., $|\psi_A\rangle\langle\psi_A| \otimes |\psi_B\rangle\langle\psi_B|$). We can interpret this as searching for the separable state that has the highest chance of being accepted by the POVM measurement $\{M, \mathsf{Id} - M\}$. For simplicity, in the rest of the paper we will specialize to the case where $d_A = d_B$; it is easy to reduce general instances of the problem to this case [HM13].

More generally, we can consider state spaces that are built of tensor products of many factors, i.e. $\mathcal{H} = (\mathcal{H}_A)^{\otimes k}$, and define the set of k-partite separable states as

$$\mathrm{Sep}^k(d_A) = \mathrm{conv}(\{\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_k : \rho_1, \ldots, \rho_k \in \mathrm{Dens}(\mathcal{H}_A)\}).$$

One can likewise consider the multipartite separability problem $h_{\mathrm{Sep}^k(d_A)}(M)$.

2.3. SoS hierarchies for quantum problems. As noted in the introduction, the problem h_{Sep} is an instance of polynomial optimization: in the bipartite case, the objective function is a degree-4 polynomial while the constraints are degree-2 polynomials in the coefficients of the state vector. It is possible to derive an SoS hierarchy for this problem by starting with this formulation and applying the procedure described in Sect. 2.1. However, for the separability problem it is more convenient to consider an equivalent formulation

couched in the language of quantum information, called the DPS hierarchy [DPS03]. The value of the k-th level of DPS is given by $\mathrm{DPS}_k(M) = \max_{\rho \in k-\mathrm{ExtPPT}} \mathrm{Tr}[\rho M]$, where the set $k-$ExtPPT is the set of "k-extendable PPT states": all bipartite states ρ that can be obtained as the *reduced* state of a k-partite state ρ', whose support lies entirely in the k-partite symmetric subspace, and whose *partial transpose* (interchange of row and column indices) over any subset of subsystems is positive semidefinite. This equivalence between this formulation and the usual formulation of SoS can be seen by interpreting the entries of ρ as the pseudo-expectations of degree-k monomials of the coefficients of the state vector; see [DPS03] or [BBH+12] for more discussion of this point.

In the context of entangled games, one can follow a similar principle and develop an SoS hierarchy for non-commutative variables to approximate the set of correlations achievable with local measurements of quantum states [NPA08, DLTW08, BFS15]. Here, the usual formulation of k-the level of ncSoS corresponds to the non-local correlations achievable by applying a degree-k monomial operator (i.e., measurement) on the shared quantum state.

3. Framework of Deriving Lower Bounds

In this section, we demonstrate our framework of deriving sum-of-squares (SoS) or semidefinite programming (SDP) lower bounds for optimization problems. To this end, we formalize the familiar notions of optimization problem, SDP extended formulations and integrality gaps. Then we show general methods for reducing optimization problems to each other as well as mapping integrality gaps for one problem/relaxation pair to another.

3.1. Optimization problems and integrality gaps. We formulate the following abstract definition of optimization problem. This definition does not address the computational difficulty of *solving* the problem, which can often be NP-hard or even uncomputable.

Definition 3.1 (*Optimization Problem*). An optimization problem A, denoted by $\Delta^A = \{\Delta_n^A\}_{n\in\mathbb{N}}$, is a family of collections of optimization instances that are parameterized by the instance size $n \in \mathbb{N}$ and which consists of following components:

- **Feasible Set** $\mathcal{P}_n^A$ is the set of feasible solutions.
- **Instances** Δ_n^A is the set of instances (or objective functions), each of which is a map $\Phi : \mathcal{P}_n^A \mapsto [0, 1]$.
- **Optimum Value** Given n and $\Phi \in \Delta_n^A$, the optimum value of the instance Φ is

$$\mathrm{OPT}(\Phi) := \max_{x \in \mathcal{P}_n^A} \Phi(x).$$

In general we will choose our parametrization so that the number of variables in the problem is equal to n or to a polynomial function of n. Note that under this defintion, the feasible set is allowed to depend only on the instance size n and not on the instance itself. An example of an optimization problem that fits under this definition is MAX-CUT, in which n is the number of vertices in a graph, $\mathcal{P}_n^{\text{MAX-CUT}} = \{0, 1\}^n$ and $\Delta_n^{\text{MAX-CUT}}$ is the set of functions of the form $\Phi(x) := \mathbb{E}_{(i,j)\sim E}(x_i - x_j)^2$ for some $E \subset [n] \times [n]$. As we can see from the example, the functions Φ can usually be efficiently specified (in this case by the edge set E), and can be thought of as the computational "question" while the optimal value of x can be thought of as the "answer."

We will focus on the following important special cases of optimization problems.

Definition 3.2 (*Polynomial Optimization*). A *polynomial* optimization problem A is an optimization problem in which the feasible set $\mathcal{P}_n^A$ is a variety of $\mathbb{R}^n$ defined by m polynomial constraints for some bounded function $m = m(n)$ and in which each instance $\Phi \in \Delta_n^A$ is a polynomial function from $\mathcal{P}_n^A$ to $[0, 1]$. Here "variety" means that

$$\mathcal{P}_n^A = \{x \in \mathbb{R}^n : g_1(x) = \cdots = g_m(x) = 0\},$$

for some polynomials $g_1, \ldots, g_m$.

Definition 3.3 (*Boolean Polynomial Optimization*). A *boolean polynomial* optimization problem A, denoted by Π^A, is a polynomial optimization problem defined by the constraints $x_i^2 = 1$ for $i = 1, \ldots, n$. Thus the feasible set $\mathcal{P}_n^A$ for such a problem is always the boolean hypercube $\mathcal{P}_n^{\text{bool}} = \{\pm 1\}^n$.

It is easy to see that the above definitions of optimization problems capture many problems of interest. For example, MAX-3-SAT, MAX-CUT and other MAX-CSPs can be formulated as boolean polynomial optimization problems with the objective function being a polynomial that counts the fraction of satisfiable clauses, as indicated below.

Definition 3.4 (*Constraint Satisfaction Problem*). A (maximum) *constraint satisfaction problem* ((MAX)-CSP) A is a type of optimization problem over the boolean hypercube $\{\pm 1\}^n$ specified by a collection of *clauses* $\{h_1, \ldots, h_m\}$, where each clause $h_i : \{\pm 1\}^n \to \{0, 1\}$ is a boolean function, and the objective function $f = \frac{1}{m} \sum_{i=1}^m h_i$ counts the fraction of clauses that evaluate to 1.

Proposition 3.5. *Any CSP where each constraint depends on $\leq \kappa$ variables can be written as a boolean polynomial optimization problem, where the objective function is a polynomial of degree κ.*

Proof. This is a simple example of polynomial interpolation. Assume for ease of notation that each constraint depends on exactly κ variables. We first show that each constraint can be expressed as a low-degree polynomial. Given a string $(y_1, \ldots, y_\kappa) \in \{\pm 1\}^\kappa$, we define the *indicator* function

$$\mathbf{1}_{y_1,\ldots,y_\kappa}(x_1, \ldots, x_\kappa) = \prod_{i=1}^{\kappa} \frac{1 + x_i y_i}{2}.$$

This function is a polynomial of degree κ. It is easy to see that when $x_i = y_i$ for all i, then $\mathbf{1}_{y_1,\ldots,y_\kappa}(x_1, \ldots, x_\kappa) = 1$, and if $x_i \neq y_i$ for any i, then $\mathbf{1}_{y_1,\ldots,y_\kappa}(x_1, \ldots, x_\kappa) = 0$. Using these indicator functions, we can express any boolean function h over κ variables as a polynomial with degree κ:

$$h(x_1, \ldots, x_\kappa) = \sum_{(y_1,\ldots y_\kappa) \in \{\pm 1\}^\kappa} h(y_1, \ldots, y_\kappa) \mathbf{1}_{y_1,\ldots,y_k}(x_1, \ldots, x_\kappa).$$

Thus, if we are given a CSP with clauses $\{h_1, \ldots, h_m\}$, each of which depends on κ variables, then the total objective function $f = \frac{1}{m} \sum_{i=1}^m h_i$ is a polynomial of degree κ. $\square$

Another important class of optimization problems are operator norms of linear functions, defined as follows. If $\mathcal{A}, \mathcal{B}$ are normed spaces and $T : \mathcal{A} \to \mathcal{B}$ is a linear map then $\|T\|_{\mathcal{A}\to\mathcal{B}} := \sup_{x\neq 0} \|T(x)\|_{\mathcal{B}}/\|x\|_{\mathcal{A}}$ can be thought of as an optimization problem where $\mathcal{P}_n$ is the unit ball of $\mathcal{A}$ and $\Phi_n^A(x) = \|T(x)\|_{\mathcal{B}}$. If $\mathcal{A} = \ell_p^n$ and $\mathcal{B} = \ell_q^m$ then this corresponds to a polynomial optimization problem. Computing h_{Sep} (cf. (1.1)) can be similarly be formulated as a polynomial optimization problem, where the feasible set is the unit sphere [BBH+12].

Our goal is to find optimization problems where the SoS hierarchy and other SDP extended formulations fail. These examples are known as "integrality gaps," where the terminology comes from the idea of approximating integer programs with convex relaxations. For our purposes, an integrality gap will be an example of an optimization problem in which the true answer is lower than the output of the SDP extended formulation. To achieve this, we need to demonstrate a feasible point of the SDP with a value that is larger than the true answer. These feasible points are called *pseudo-solutions*, and we will define them for any polynomial optimization problem as follows.

Definition 3.6 (*Pseudo-Solution*). Let A be a polynomial optimization problem. Let $\Phi_n^A \in \Delta_n^A$ be an instance of optimization A for some n. A *degree-d value-c pseudo-solution* for Φ_n^A is a degree-d pseudo-expectation $\tilde{\mathbb{E}}$ satisfying the constraints of $\mathcal{P}_n^A$ such that

$$\tilde{\mathbb{E}}[\Phi_n^A(x)] \geq c.$$

In the case of CSPs, we can also define a stronger type of pseudo-solution that not only achieves an objective value of 1, but also satisfies the constraints of the CSP as "hard" constraints.

Definition 3.7. Given a CSP A whose objective function is a polynomial of degree k, we say a degree-d pseudo-expectation $\tilde{\mathbb{E}}[\cdot]$ *perfectly satisfies* A if for every constraint $g_i(x)$ of A, and every polynomial $p(x)$ with $\deg(p) \leq d - k$,

$$\tilde{\mathbb{E}}[p(x)(g_i(x) - 1)] = 0.$$

A single degree-d value-c pseudo-solution for an instance Φ_n^A implies that the sum-of-squares approach (up to degree d) believes the optimum value of Φ_n^A is at least c. If the true optimum value of Φ_n^A is smaller than c, then such a pseudo-solution serves as an integrality gap for the SoS approach, i.e. an example where the SoS hierarchy gives the wrong answer. To refute the power of the SoS hierarchy, we need to establish such pseudo-solutions as well as small true optimum values for any large n.

Definition 3.8 (*Integrality gap*). Let A be any polynomial optimization problem. Let $d = d(n), c = c(n), s = s(n)$ be functions of n such that $0 \leq s < c \leq 1$. A degree-d value-(c, s) integrality gap for A is a collection of $\Phi_n^A \in \Delta_n^A$ for each $n \geq n_0$, s.t.

- The true optimum value $\mathrm{OPT}(\Phi_n^A) \leq s$.
- For each $n \geq n_0$, there exists a degree-d value-c pseudo-expectation $\tilde{\mathbb{E}}_n$ for Φ_n^A such that $\tilde{\mathbb{E}}_n[\Phi_n^A(x)] \geq c$.

We can relate integrality gaps to lower bounds on $\deg_{\mathsf{SOS}}$ as follows.

Proposition 3.9. *Let A be any polynomial optimization problem with a degree-d value-(c, s) integrality gap. For some $0 < \delta \leq c - s$ let $f_n = c - \delta - \Phi_n^A(x_n^A)$ where Φ_n^A is from the integrality gap and $x_n^A \in \mathcal{P}_n^A$. Then f_n is a polynomial taking nonnegative values over $\mathcal{P}_n^A$ and has $\deg_{\mathsf{SOS}}(f_n) > d$.*

Proof. Immediate from the definitions. □

3.2. Reduction between optimization problems. To obtain SoS lower bounds for optimization problems, it suffices to establish integrality gaps. However, it is not clear how to obtain such integrality gaps in general, which might be a challenging task on its own. Here, we formulate an approach to establish such integrality gaps through reductions. Specifically, we start with some optimization problem with known integrality gaps and reduce it to an optimization problem that we want to establish integrality gaps.

Definition 3.10. (*Reductions*) A *reduction* $R_{A \Rightarrow B}$ from optimization problem A to optimization problem B is a map from Δ^A to Δ^B; i.e. $R(\Phi_{n_A}^A) \in \Delta_{n_B}^B$, where n_B is a function of n_A.

We remark that for the purpose of establishing integrality gaps, the reduction needs to be neither explicit or efficient. However, it is favorable to have the following properties for the reduction.

Definition 3.11. (*Properties of Reductions*) A *reduction* $R_{A \Rightarrow B}$ from optimization problem A to optimization problem B is called

- (s^B, s^A)**-approximate**[3] if for any n and any $\Phi_{n_A}^A$ and its corresponding $\Phi_{n_B}^B = R(\Phi_{n_A}^A)$, we have

$$\mathrm{OPT}(\Phi_{n_A}^A) = \max_{x \in \mathcal{P}_{n_A}^A} \Phi_{n_A}^A(x) \leq s^A \Rightarrow \mathrm{OPT}(\Phi_{n_B}^B) = \max_{x \in \mathcal{P}_{n_B}^B} \Phi_{n_B}^B(x) \leq s^B.$$

  Here s^A, s^B are understood to be functions of n_A, n_B, and $s^A \leq s^B$.
- **embedded** if for any n_A, there is an additional map $E : \mathcal{P}_{n_A}^A \mapsto \mathcal{P}_{n_B}^B$ such that for any $\Phi_{n_A}^A$ and its corresponding $\Phi_{n_B}^B$, any $x_{n_A}^A \in \mathcal{P}_{n_A}^A$ and its corresponding $x_{n_B}^B = E(x_{n_A}^A) \in \mathcal{P}_{n_B}^B$, we have

$$\Phi_{n_A}^A(x_{n_A}^A) = \Phi_{n_B}^B(x_{n_B}^B).$$

The first property shows the soundness of the reduction, while the second property can be viewed as a strong statement about completeness. Not only should the optimum value of the reduced problem be at least as large, but each $x \in \mathcal{P}_n^A$ (i.e. including non-optimal x) corresponds to some point in $\mathcal{P}_n^B$ with the same value under Φ_n^B. This condition was needed for the recent SDP lower bounds in [LRS15].

For reductions between polynomial optimization problems, the following property is crucial to establish pseudo-solutions, and eventually integrality gaps, for the reduced problems.

[3] We write the parameters in this order in order to match the convention (c, s)-approximate extended formulations, which are defined below. We will see that a (s^B, s^A)-approximate reduction gives rise to a $(c = s^B, s = s^A)$-approximate extended formulation.

Definition 3.12. (*Pseudo-solution Preserving Reduction*) Let $R_{A\Rightarrow B}$ be a reduction from polynomial optimization problem A to polynomial optimization problem B. It is called (d^A, c^A, d^B, c^B) **pseudo-solution preserving** if for any degree-d^A value-c^A pseudo-solution for any instance $\Phi^A_{n_A}$, there is a degree-d^B value-c^B pseudo-solution for its corresponding instance $\Phi^B_{n_B}$, for any n_A. Here d^A, c^A, d^B, c^B should be thought of as functions of n_A.

It is straightforward to verify that the above three properties are transitive. Thus, it is possible to design a chain of reductions for complicated reductions.

Proposition 3.13. (Transitivity of Properties of Reductions) *Let $R_{A\Rightarrow B}$ and $R_{B\Rightarrow C}$ be reductions from optimization problem A to optimization problem B and from optimization problem B to optimization problem C respectively. Let $R_{A\Rightarrow C}$ be the natural composition of $R_{A\Rightarrow B}$ and $R_{B\Rightarrow C}$.*

- *If $R_{A\Rightarrow B}$ is (s^B, s^A)-approximate and $R_{B\Rightarrow C}$ is (s^C, s^B)-approximate, then $R_{A\Rightarrow C}$ is (s^C, s^A)-approximate.*
- *If $R_{A\Rightarrow B}$ and $R_{B\Rightarrow C}$ are embedded, then $R_{A\Rightarrow C}$ is embedded.*
- *If $R_{A\Rightarrow B}$ is (d^A, c^A, d^B, c^B) pseudo-solution preserving and $R_{B\Rightarrow C}$ is (d^B, c^B, d^C, c^C) pseudo-solution preserving, then $R_{A\Rightarrow C}$ is (d^A, c^A, d^C, c^C) pseudo-solution preserving.*

Proof. Immediate from the definitions. □

We are ready to illustrate how reductions help establish integrality gaps.

Proposition 3.14. *Let A, B be polynomial optimization problems. Let $R_{A\Rightarrow B}$ be the reduction from A to B. Assuming there exists a degree-d^A value-(c^A, s^A) integrality gap for A, if $R_{A\Rightarrow B}$ is (s^B, s^A)-approximate and (d^A, c^A, d^B, c^B) pseudo-solution preserving, then there exists a degree-d^B value-(c^B, s^B) integrality gap for B.*

Proof. It follows directly by definition. □

As a direct consequence, Proposition 3.14 suggests that we can make use of reductions to derive SoS lower bounds. The hard part is, however, the design of reductions with approximation and pseudo-solution preserving properties. Here, we describe a simple but useful observation that constructs a pseudo-solution from another under a polynomial map. We say that $p : \mathbb{R}^n \to \mathbb{R}^m$ is a degree-d polynomial map if $p(x) = (p_1(x), \ldots, p_m(x))$ where each $p_i \in \mathbb{R}[x]_d$.

Lemma 3.15. *Let $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$ be algebraic varieties, meaning that*

$$A = \{x \in \mathbb{R}^n : g_1(x) = \cdots = g_{n'}(x) = 0\} \tag{3.1a}$$

$$B = \{x \in \mathbb{R}^m : h_1(x) = \cdots = h_{m'}(x) = 0\}, \tag{3.1b}$$

for some polynomials $\{g_i\}, \{h_i\}$.

Suppose that p is a degree-d polynomial map from $\mathbb{R}^n \to \mathbb{R}^m$ such that $p(A) \subseteq B$. Let $\tilde{\mathbb{E}}_A \in \mathbb{R}[x_1, \ldots, x_n]^*_\ell$ be a degree-ℓ pseudo-expectation that is compatible with the constraints $g_1, \ldots, g_{n'} \in \mathbb{R}[x_1, \ldots, x_n]$. Then there exists a degree-ℓ/d pseudo-expectation $\tilde{\mathbb{E}}_B \in \mathbb{R}[y_1, \ldots, y_m]^*_{\ell/d}$ that is compatible with the constraints $h_1, \ldots, h_{m'}$.

Note that this is essentially the same statement as Fact A.8 in [BKS14], which was stated there without proof.

Proof. Assume that A is a discrete set. This matches our actual application in which $A = \{\pm 1\}^n$ and mostly affects only the notation. We will need to introduce the notion of a pseudo-density. A degree-l pseudo-density on A is a function $\mu_A : A \to \mathbb{R}$ such that $\sum_{x\in A} \mu_A(x) = 1$ and $\sum_{x\in A} \mu_A(x) f(x)^2 \geq 0$ for all $f \in \mathbb{R}[x]_{l/2}$. The term "pseudo-" refers to the fact that $\mu_A(x)$ can be negative. Any true probability distribution is also a pseudo-density and in the case of $A = \{\pm 1\}^n$, degree-n pseudo-densities are also probability distributions. In general a degree-ℓ pseudo-density μ_A induces a pseudo-expectation $\tilde{\mathbb{E}}_A \in \mathbb{R}[x]_\ell^*$ with

$$\tilde{\mathbb{E}}_A[f] := \sum_{x\in A} \mu_A(x) f(x), \tag{3.2}$$

for all $f \in \mathbb{R}[x]_l$.

To obtain a pseudo-density from a pseudo-expectation we need to solve an underconstrained system of linear equations. This can be done as follows. Let $e_A : \mathbb{R}[x] \to \mathbb{R}^A$ denote the evaluation map on A; i.e. $e_A(f)$ is the tuple $(f(x))_{x\in A}$. Note that e_A is a linear map, and we can also view μ_A as a linear map from $\mathbb{R}^A \to \mathbb{R}$. Given a degree-$\ell$ pseudo-expectation $\tilde{\mathbb{E}}_A$, (3.2) can be thought of as constraining μ_A on the subspace $e_A(\mathbb{R}[x]_\ell)$. If we write $\mathbb{R}^A = e_A(\mathbb{R}[x]_\ell) \oplus V$ for some subspace V then we can extend μ_A to act arbitrarily on V. As long as the action on $\mathbb{R}[x]_\ell$ is the same, this will still meet the definition of a pseudo-distribution.

Now define

$$\mu_B(y) = \sum_{x\in p^{-1}(y)} \mu_A(x), \quad \forall y \in B.$$

Since $p(A) \subseteq B$ we have $\sum_{y\in B} \mu_B(y) = \sum_{x\in A} \mu_A(x) = 1$. And if $f \in \mathbb{R}[y_1, \ldots, y_m]_{\ell/2d}$ then

$$\sum_{y\in B} \mu_B(y) f(y)^2 = \sum_{x\in A} \mu_A(x) f(p(x))^2 \geq 0, \tag{3.3}$$

since $\deg(f \circ p) \leq \ell/2$. Thus μ_B is a valid pseudo-density.

Finally, we can define $\tilde{\mathbb{E}}_B \in \mathbb{R}[y]_{\ell/d}^*$ by

$$\tilde{\mathbb{E}}_B[f] := \sum_{y\in B} \mu_B(y) f(y). \tag{3.4}$$

By the above arguments, $\tilde{\mathbb{E}}_B[1] = 1$ and $\tilde{\mathbb{E}}_B[f^2] \geq 0$ whenever $\deg f \leq \ell/2d$. Also, for any $i \in [m']$ and any $q \in \mathbb{R}[y]_{\ell/d-\deg(h_i)}$ we have

$$\tilde{\mathbb{E}}_B[h_i q] = \sum_{y\in B} \mu_B(y) h_i(y) q(y) = 0, \tag{3.5}$$

since $h_i(B) = 0$. Thus $\tilde{\mathbb{E}}_B$ is compatible with the constraints $h_1, \ldots, h_{m'}$. □

The previous lemma implies the following corollary, which allows us to obtain perfectly satisfying pseudo-solutions for CSPs via "local" reductions.

Proposition 3.16. *Let A, B be CSPs with a reduction $R_{A\Rightarrow B}$. Suppose that*

- *there exists a map* $f : \mathcal{P}^A_{n_A} \to \mathcal{P}^B_{n_B}$*, such that if* $x_A \in \mathcal{P}^A_{n_A}$ *satisfies all the constraints of an instance of* A*, then* $f(x_A) \in \mathcal{P}^B_{n_B}$ *satisfies all the constraints of the corresponding instance of* B*,*
- *each coordinate of* $f(x_A)$ *depends on at most* κ *coordinates of* x_A*, and*
- *there exists a degree-*d^A *pseudo-solution that perfectly satisfies* A*.*

Then there exists a degree-d^A/κ *pseudo-solution that perfectly satisfies* B*.*

Proof. By a similar procedure to the proof of Proposition 3.5, we can express each coordinate of the mapping f as a polynomial of degree κ. Now, suppose that the source problem A has a perfectly satisfying degree-d^A pseudo-solution, given by a pseudo-expectation operator $\tilde{\mathbb{E}}^A[\cdot]$. Then Lemma 3.15 applied to the degree-κ polynomial maps constructed above, yields a degree-d^A/κ perfectly satisfying pseudo-solution for B. □

3.3. 3XOR *with integrality gap.* In this section, we will introduce the base hard problem underlying our reductions, which is the 3XOR problem first discovered by Grigoriev [Gri01] and subsequently rediscovered by Schoenebeck [Sch08]. It is analogous to the proof that 3-SAT is NP-hard, from which other hardness results can be derived by reducing those problems to 3-SAT. In our framework, 3XOR can be formulated as follows.

Definition 3.17. (3XOR) 3XOR is a boolean polynomial optimization problem with the following restriction:

- **Instances** for any n, an instance is parameterized by a formula Φ_n that consists of a set $\mathcal{C}$ of $m = m(n)$ 3XOR clauses on n boolean (± 1) variables. In other words, we have the constraints $x_i^2 = 1$ for each i and the objective function is

$$\Phi_n(x) = \frac{1}{m} \sum_{(i,j,k)\in\mathcal{C}} \frac{1 + a_{ijk} x_i x_j x_k}{2}.$$

Thanks to the $x_i^2 = 1$ constraints, these terms are equivalent to ones of the form $(1 - (x_i x_j x_k - a_{ijk})^2)/2$.

Grigoriev's result [Gri01] (reformulated by Barak [Bar14]) implies the following integrality gaps. We have a slightly different formulation from [Bar14] that is slightly stronger but guaranteed by [Gri01].

Proposition 3.18. (Theorem 3.1 of [Bar14], due to Grigoriev) *For any* $\epsilon > 0$*, for every* n *large enough there exists a* 3XOR *instance* Φ_n *with* n *variables and* $m = O(n/\epsilon^2)$ *clauses, such that* $\mathrm{OPT}(\Phi_n) \le \frac{1}{2} + \epsilon$*, but there exists a degree-*$\Omega(n)$ perfectly satisfying *pseudo-solution* $\tilde{\mathbb{E}}$*. In other words there is a degree-*$\Omega(n)$ *value-*$(1, \frac{1}{2} + \epsilon)$ *integrality gap for* 3XOR.

Recall that "perfectly satisfying" means that for every clause $x_i x_j x_k = a_{ijk}$, it holds that $\tilde{\mathbb{E}}[(x_i x_j x_k - a_{ijk}) p(x)] = 0$ for all polynomials $p(x)$ with degree at most $d - 3$.

4. Lower Bounds on h_{Sep} and Its Applications

In this section, we will explain how the lower bounds on h_{Sep} are derived through reductions in the framework introduced in Sect. 3. We will describe the high-level reduction path here and then explain each reduction in detail in following subsections.

$$\text{3XOR} \underset{R_1}{\Longrightarrow} \text{2- OUT- OF- 4- SAT- EQ} \underset{R_2}{\Longrightarrow} \text{QMA(2)- HONEST} \underset{R_3}{\Longrightarrow} h_{\text{Sep}}.$$

There are three reductions R_1, R_2, R_3 respectively in the reduction path from 3XOR to h_{Sep}. The starting point is 3XOR as we introduced in Sect. 3.3. We need to define two intermediate problems.

- 2- OUT- OF- 4- SAT- EQ is a boolean polynomial optimization in which each instance is parameterized by a formula Φ_n that consists of 2- OUT- OF- 4 clauses and EQ clauses.
  - Each 2- OUT- OF- 4 clause involves 4 boolean variables $x_i, x_j, x_k, x_l \in \{\pm 1\}$. The clause is satisfied if and only if exactly 2 out of 4 variables x_i, x_j, x_k, x_l are true (+1).
  - Each EQ clause involves 2 boolean variable $x_a, x_b \in \{\pm 1\}$. The clause is satisfied if and only if $x_a = x_b$.

  These acceptance conditions correspond to the 0/1-valued predicates

$$\frac{1 + x_i x_j x_k x_l}{2} \quad \text{and} \quad \frac{1 + x_a x_b}{2}. \tag{4.1}$$

  If we define $\Phi_n(x)$ to be the fraction of clauses in Φ_n satisfied by x we can then see that $\Phi_n(x)$ is a degree-4 polynomial in $x_1, \ldots, x_n$.
- QMA(2)- HONEST is the "honest prover" acceptance probability of a QMA(2) protocol (see Sect. 4.2 for details) for 2- OUT- OF- 4- SAT- EQ This QMA(2) protocol

We note that most of this chain of reductions is implicit in the earlier works of Aaronson, Beigi, Drucker, Fefferman and Shor (ABDFS) [ABD+08] and Harrow and Montanaro [HM13], which show reductions from 3SAT to h_{Sep}. The only exception is that we replace the application of PCP theorem, which used to be the first step in reductions and turns out to be a high-degree reduction, by some direct and low-degree construction inspired by part of the proof of the PCP theorem.

Moreover, our argument requires explicit analysis of the intermediate steps of the chain to make sure that individual reductions are pseudo-solution preserving (see Definition 3.12) and thus low-degree. These explicit analysis of each step will help us further to enable application of the LRS result, which we will elaborate on in Sect. 5.

Precise definitions of each problem will appear in each corresponding subsection. All three reductions will be elaborated on in Sects. 4.1 and 4.2, as well as the SoS hardness result of the h_{Sep} and 2-to-4 norm problem. We will briefly describe extensions to other ETH-based hardness results in Sect. 4.3.

4.1. From 3XOR *to* 2- OUT- OF- 4- SAT- EQ. In this section we show an explicit reduction from 3XOR to 2- OUT- OF- 4- SAT- EQ that preserves the pseudo-solutions and has reasonable approximation parameters. The following proposition shows the reduction has reasonable approximation parameters as well as some other useful features for later reduction steps.

Proposition 4.1. *For all m, n, there exists a reduction that maps a given* 3XOR *instance Ψ with m clauses and n variables, onto a* 2- OUT- OF- 4- SAT- EQ *instance Φ satisfying the following properties:*

1. *Every variable in Φ appears in at most $O(1)$ clauses.*
2. *Φ has $O(n+m)$ variables and $O(m)$ clauses.*
3. *If Ψ is perfectly satisfiable, then so is Φ.*
4. *If at most $1-\delta$ fraction of the clauses of Ψ are satisfiable, then at most $1-\Omega(\delta)$ fraction of the clauses of Φ are satisfiable.*

Proof. We perform the reduction in two steps. First, we reduce the 3XOR instance to a 2- OUT- OF- 4- SAT- EQ instance in a manner that preserves properties (4.1)-(4.1). Next, we achieve property (4.1) without losing the others through an "expanderizing" step, similar to the degree reduction in Dinur's proof of the PCP theorem. We now describe the two steps in turn.

Step 1 We show how to transform each 3XOR clause into three 2- OUT- OF- 4 clauses, each acting on one of the original 3XOR variables and two new dummy variables. Altogether we introduce three new dummy variables per 3XOR clause. Additionally, in order to break the symmetry of 2- OUT- OF- 4- SAT under parity reversal, we introduce a *parity reference bit*, which we denote z. Suppose for now that $z = 1$. Now, suppose we have a 3XOR clause $c = [x_i x_j x_k = a_{ijk}]$. First we treat the case where $a_{ijk} = 1$. We introduce three new variables y_1^c, y_2^c, y_3^c, and generate the following three 2- OUT- OF- 4 clauses: 2- OUT- OF- $4(x_i, y_2^c, y_3^c, z)$, 2- OUT- OF- $4(x_j, y_1^c, y_3^c, z)$, and 2- OUT- OF- $4(x_k, y_1^c, y_2^c, z)$. If we fix an assignment to x_i, x_j, x_k, it is easy to see that if $x_i x_j x_k = a_{ijk}$ then there exists an assignment to y_1^c, y_2^c, y_3^c that satisfies all three clauses; otherwise, at most two of the three clauses are satisfied for all assignments to y_1^c, y_2^c, y_3^c. In particular, if $(x_i, x_j, x_k) = (1, 1, 1)$, then we set $(y_1^c, y_2^c, y_3^c) = (-1, -1, -1)$ and if (x_i, x_j, x_k) has Hamming weight 1 then we set $(y_1^c, y_2^c, y_3^c) = (x_i, x_j, x_k)$. On the other hand, since each y_i^c appears in two clauses, multiplying all the clauses yields $(x_i y_2^c y_3^c z)(x_j y_1^c y_3^c z)(x_k y_1^c y_2^c z) = x_i x_j x_k z$. If this equals -1, then not all of the 2- OUT- OF- 4 clauses can be satisfied. If $a_{ijk} = -1$, then we simply replace z with $\neg z$ in the 2- OUT- OF- 4 clauses, and the same story holds.

Applying this transformation to all the clauses of the 3XOR instance yields a 2- OUT- OF- 4- SAT instance with $n + 3m + 1$ variables and $3m$ clauses. If the original satisfying fraction was 1, then the resulting instance also has satisfying fraction 1; otherwise, if the original satisfying fraction was $1-\delta$, the new instance has satisfying fraction at most $1-\delta/3$.

The above analysis holds only when $z = 1$. If we set $z = -1$ then all satisfied 3XOR clauses become unsatisfied and vice-versa. However, this symmetry already existed in the original 3XOR formula. Indeed replacing $x_1, \ldots, x_n$ with $-x_1, \ldots, -x_n$ would have the same effect. Thus we can assume WLOG that $z = 1$.

Step 2 The resulting 2- OUT- OF- 4- SAT instance may have some variables that occur in a large number of clauses. Indeed, the parity reference bit occurs in all of the clauses. To fix this, we shall apply Lemma 4.2 that fixes this issue while keeping all other properties. □

Lemma 4.2. (Degree reduction) *There exists a process that maps any instance G of* 2- OUT- OF- 4- SAT *to an instance G' of* 2- OUT- OF- 4- SAT- EQ *where*

1. *Every variable appears in at most 4 constraints.*

2. *If G has m clauses, then G' has $\leq O(m)$ clauses.*
3. *If* $\mathrm{OPT}(G) = 1$, *then* $\mathrm{OPT}(G') = 1$.
4. *If* $\mathrm{OPT}(G) = 1 - \epsilon$, *then* $\mathrm{OPT}(G') \leq 1 - \eta\epsilon$ *for constant* η.

Proof. We use the "expanderization" process introduced by Papadimitriou and Yannakakis [PY91]. Specifically, we replace every variable that occurs in too many clauses by copies, with equality checks between them arranged according to a degree-3 expander graph. □

In the following we demonstrate the above reduction also preserves the pseudo-solutions.

Proposition 4.3. *For some constant* $0 < \delta < 1$, *there exists a degree-*$\Omega(n)$ *value-*$(1, 1-\delta)$ *integrality gap for the* 2- OUT- OF- 4- SAT- EQ *problem. Moreover, if for any* 2- OUT- OF- 4- SAT *clause* 2- OUT- OF- $4(x_i, x_j, x_k, x_\ell)$ *in any instance* Φ_n, *we have for any polynomial* $p(x)$ *of degree at most* $d - 4$,

$$\tilde{\mathbb{E}}[p(x)(x_i + x_j + x_k + x_\ell)] = 0,$$

where $\tilde{\mathbb{E}}$ *is from the pseudo-solution of the integrality gap.*

To keep the notation simple in the above proposition we ignore the fact that a 2- OUT- OF- 4 clause could in general be of the form 2- OUT- OF- $4((-1)^a x_i, (-1)^b x_j, (-1)^c x_k, (-1)^d x_\ell)$.

Proof. We start with the degree-$\Omega(n)$ value-$(1, \frac{1}{2} + \epsilon)$ integrality gap from Proposition 3.18 for 3XOR. Using the reduction in Proposition 4.1, we obtain corresponding instances in the 2- OUT- OF- 4- SAT- EQ problem that have true optimum value at most $1 - \delta$ for some constant $0 < \delta < 1$.

It then suffices to establish pseudo-solutions for these instances in the 2- OUT- OF- 4- SAT- EQ problem. To do this, we recall the map between satisfying assignments defined in Proposition 4.1:

- Each variable $x \in \{\pm 1\}$ from the original 3XOR instance is mapped to a variable in the 2- OUT- OF- 4- SAT- EQ instance with the same assigned value.
- The 2- OUT- OF- 4- SAT- EQ instance has a parity reference bit z that is set to be 1.
- For each 3XOR clause c, we introduce 3 dummy variables y_1^c, y_2^c, y_3^c in the 2- OUT- OF- 4- SAT- EQ instance. For every satisfying assignment to the clause c, there exists an satisfying assignment to the dummy variables that depends only on the assignments of the variables in the clause c.
- The copies of variables in the expanderization step (Lemma 4.2).

Thus, the hypotheses of Proposition 3.16 are satisfied with $\kappa = 3$. Hence, by applying that proposition to the perfectly satisfying pseudo-solution given by Proposition 3.18, we obtain a degree-$\Omega(n)$ perfectly satisfying pseudo-solution as desired. All in all, this gives us a degree-$\Omega(n)$ value-$(1, 1-\delta)$ integrality gap for 2- OUT- OF- 4- SAT- EQ. □

4.2. From 2- OUT- OF- 4- SAT- EQ *to* QMA(2)- HONEST. We start with a description of a QMA(2) protocol for the problem 2- OUT- OF- 4- SAT- EQ, and the formal definition of QMA(2)- HONEST. We will show that the fact that our QMA(2) protocol solves 2- OUT- OF- 4- SAT- EQ implies a reduction from 2- OUT- OF- 4- SAT- EQ reduces to QMA(2)- HONEST. Later on, we will use the connection between QMA(2) and h_{Sep} to reduce from QMA(2)- HONEST to h_{Sep}.

Our protocol is constructed by starting with a slight modification of the QMA(k) protocol due to ABDFS [ABD+08] and then converting this QMA(k) protocol into a QMA(2) protocol using reduction of Harrow and Montanaro [HM13]. The following is a precise statement of the properties of the resulting QMA(2) protocol.

Proposition 4.4. *For any constant* $0 < \delta < 1$ *and any constant* $\epsilon > 0$*, there exists a QMA(2) protocol P for* 2- OUT- OF- 4- SAT- EQ *such that*

1. *If any instance* Φ *of* 2- OUT- OF- 4- SAT- EQ *has* $\mathrm{OPT}(\Phi) = 1$*, then protocol P accepts with probability 1 with the following quantum witness* $|\psi_x\rangle \otimes |\psi_x\rangle$*, where*[4]

$$|\psi_x\rangle = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i\,|i\rangle\right)^{\otimes \tilde{O}(\sqrt{n})}, \forall \textit{ satisfiable assignment } x \in \{\pm 1\}^n. \tag{4.2}$$

2. *If any instance* Φ *of* 2- OUT- OF- 4- SAT- EQ *has* $\mathrm{OPT}(\Phi) \leq 1 - \delta$*, then protocol P accepts with probability at most* ϵ *on any separable quantum witness.*

Proof. We construct such a QMA(2) protocol by composition of a slight modification of the QMA(k) protocol due to [ABD+08] and [HM13].

The QMA(k) protocol due to ABDFS [ABD+08] can be used to solve 2- OUT- OF- 4- SAT when $k = \tilde{O}(\sqrt{n})$. So we just need to show how to modify the protocol to handle equality clauses. In the original ABDFS protocol one of the tests was to project onto a random subspace spanned by $\{|i\rangle, |j\rangle, |k\rangle, |\ell\rangle\}$ with i, j, k, ℓ corresponding to some 2- OUT- OF- 4 clause. With probability 1/2 we will replace this with a test that first projects onto the span of $\{|i\rangle, |j\rangle\}$ where i, j come from a random equality clause $x_i = x_j$. Next, we check whether the state is orthogonal to $|i\rangle - |j\rangle$. It is not hard to see the analysis therein works with this slight change, as long as we have the same regularity condition that each variable participates in $O(1)$ equality constraints.

Then we can apply the generic reduction from any QMA(k) to QMA(2) from [HM13]. The final protocol is the composition of the above two steps. It takes as input a two-party separable state, where each half of the state consists of $\sqrt{n}\,\mathrm{polylog}(n)$ qubits. These are grouped into registers of $k = \lceil \log_2(n) \rceil$ qubits each. The protocol consists of the following four tests, which we describe briefly (for full descriptions, see the orignal works):

1. Product test [HM13]: Think of the proofs from the two provers as divided into pieces $A_1, \ldots, A_m$ and $B_1, \ldots, B_m$ respectively, of k qubits each, where $m = \sqrt{n}\mathrm{polylog}(n)$. In this test the verifier projects the registers $A_i B_i$ onto the bipartite symmetric subspace $\mathrm{Sym}^2\,\mathbb{C}^{2^k}$ for each i and rejects if any $A_i B_i$ is found in the antisymmetric subspace.
2. Symmetry test [ABD+08]: In this test, the verifier projects $A_1, \ldots, A_m$ onto the m-partite symmetric subspace $\mathrm{Sym}^m\,\mathbb{C}^{2^k}$ and similarly for $B_1, \ldots, B_m$. (We could also

[4] We use the convention that $\tilde{O}(\cdot)$ hides constants as well as $\mathrm{polylog}(n)$ terms.

combine this with the Product test and simply project $A_1, \ldots, B_m$ onto $\mathrm{Sym}^{2m}\,\mathbb{C}^{2^k}$; this would leave completeness the same and would only improve the soundness. The only reason to consider the Product and Symmetry tests separately is so that we can use the analyses of [ABD+08] and [HM13] as black boxes.)

3. Uniformity test [ABD+08]: The verifier chooses a maximal matching $\mathcal{M}$ on the set $\{1, \ldots, n\}$ at random, and constructs an orthonormal basis containing the vectors
$$\frac{1}{\sqrt{2}}(|u\rangle \pm |v\rangle)$$
for every edge $(u, v) \in \mathcal{M}$. The verifier then measures each A_i and each B_j in this basis, and rejects if for some $(u, v) \in \mathcal{M}$, the outcomes $\frac{1}{\sqrt{2}}(|u\rangle \pm |v\rangle)$ are both obtained on different subsystems.
4. Satisfiability test [ABD+08], modified to handle equality clauses as described above: The verifier divides the clauses into a constant number of blocks such within each block, no two clauses share a variable. For each piece of the proof A_i or B_j, the verifier chooses a block B at random, and applies a POVM whose elements are the projectors Π_C onto the span $\mathrm{span}(\{|i\rangle : i \in C\})$ of computational basis states corresponding to the variables in each clause C in the block B. This is completed into a POVM by including the POVM element $M_{\text{default}} = I - \sum_{(i,j,k,\ell)\in B} \Pi_{i,j,k,\ell}$. If the measurement returns outcome C corresponding to some clause C, then the verifier checks whether this clause is satisfied by the witness. If the clause C is a 2- OUT- OF- 4- SAT clause on variables i, j, k, ℓ, this is done by checking whether the post-measurement state is orthogonal to $\frac{1}{2}(|i\rangle + |j\rangle + |k\rangle + |\ell\rangle)$, and if it is an equality clause on variables i, j, we check whether the post-measurement state is orthogonal to $\frac{1}{\sqrt{2}}(|i\rangle - |j\rangle)$. If ever any clause is discovered to be violated, the verifier rejects; otherwise, he accepts.

By composing the completeness and soundness results of [ABD+08] and [HM13], we obtain that the protocol given above has completeness 1 and constant soundness $s < 1$.

Finally, to achieve arbitrarily small constant soundness, we perform an amplification procedure. Since the measurement operator corresponding to an accepting outcome is separable, by Lemma 7 of [HM13], we can amplify the soundness of the protocol by performing parallel repetition: if we start with soundness s and repeat ℓ times in parallel, the new soundness is at most s^ℓ. □

We are ready to formally define QMA(2)- HONEST.

Definition 4.5. QMA(2)- HONEST is a boolean polynomial optimization problem with objective function $\Psi : \{\pm 1\}^n \to [0, 1]$ defined as follows. Let Φ be an instance of 2- OUT- OF- 4- SAT- EQ and P the corresponding protocol from Proposition 4.4. Then
$$\begin{aligned}\Psi(x) &:= \mathbf{Pr}[P \text{ accepts on input } \Phi \text{ and witness } |\psi_x\rangle \otimes |\psi_x\rangle],\\ &= \mathrm{tr}(M_n^\Phi\, |\psi_x\rangle\langle\psi_x| \otimes |\psi_x\rangle\langle\psi_x|),\end{aligned}$$
where $x \in \{\pm 1\}^n$, M_n^Φ is the POVM corresponding to the acceptance in protocol P on input Φ and $|\psi_x\rangle$ is defined in (4.2).

It is important to note that the objective function $\Psi(x)$ is a function of the boolean variables x, *not* of the components of the witness state $|\psi_x\rangle$, and hence QMA(2)- HONEST is indeed a boolean optimization problem. For this to work, we rely on the fact that $\Psi(x)$ is a degree-$\tilde{O}(\sqrt{n})$ polynomial in the variables x, which follows from the fact that the components of the witness state $|\psi_x\rangle$ are themselves degree-$\tilde{O}(\sqrt{n})$ polynomials of x.

Integrality gap for QMA(2)- HONEST We are ready to establish the integrality gap for QMA(2)- HONEST from the integrality gap for 2- OUT- OF- 4- SAT- EQ using the above reduction, i.e., each instance $\Phi \in$ 2- OUT- OF- 4- SAT- EQ is reduced to an instance $\Psi \in$ QMA(2)- HONEST as in the Definition 4.5. First, we note a useful fact that allows us to interpret pseudoexpectations as quantum states.

Definition 4.6. Let $\tilde{\mathbb{E}}_x[\cdot]$ be the pseudoexpectation operator corresponding to a degree-$2d$ pseudodistribution over n variables $\{x_1, \ldots, x_n\}$ satisfying the constraints $x_i^2 - 1 = 0$ for all i from 1 to n. (This means $\tilde{\mathbb{E}}[(x_i^2 - 1)p(x)] = 0$ for all i and for all polynomials $p(x)$ with degree at most $2d - 2$.) Then we define the associated mixed quantum state ρ_x to be the following state over $d \log n$ qubits:

$$\rho_x = \frac{1}{n^d} \sum_{\substack{i_1 i_2 \ldots i_d \\ j_1 j_2 \ldots j_d}} \tilde{\mathbb{E}}_x[x_{i_1} \ldots x_{i_d} x_{j_1} \ldots x_{j_d}] \, |i_1 \ldots i_d\rangle \langle j_1 \ldots j_d| \, .$$

If $\tilde{\mathbb{E}}_x[\cdot]$ is a valid pseudoexpectation operator, then ρ_x is PSD and has trace 1, so it is indeed a quantum mixed state. Moreover, by the $x_i^2 - 1 = 0$ constraint, we find that the reduced density matrices correspond to the low-degree moments of the pseudodistribution.

Proposition 4.7. *For any constant $\epsilon > 0$, there exists a degree-$\Omega(n)$ value-$(1, \epsilon)$ integrality gap for* QMA(2)- HONEST.

Proof. For any n, let $\Phi_n \in$ 2- OUT- OF- 4- SAT- EQ be the instance from the degree-$\Omega(n)$ value-$(1, 1 - \delta)$ integrality gap for 2- OUT- OF- 4- SAT- EQ from Proposition 4.3 and μ_n be the corresponding pseudo-solution (δ is the constant therein). Let Ψ_n be its reduced instance of QMA(2)- HONEST. For any constant $\epsilon > 0$, by Property (2) of Proposition 4.4 and the fact $\mathrm{OPT}(\Phi_n) \leq 1 - \delta$, we have $\mathrm{OPT}(\Psi_n) \leq \epsilon$. Then it suffices to establish a $(\Omega(n), 1)$ pseudo-solution for Ψ_n.

To that end, we claim that μ_n is also a degree-$\Omega(n)$ value-1 pseudo-solution for Ψ_n. Note that the feasible sets for Φ_n and Ψ_n are the same. It suffices to show $\tilde{\mathbb{E}}_{x \sim \mu}[\Psi_n(x)] = 1$. Observe that by linearity of $\tilde{\mathbb{E}}[\cdot]$ and $\mathrm{tr}(\cdot)$, we have

$$\tilde{\mathbb{E}}_{x \sim \mu}[\Psi_n(x)] = \tilde{\mathbb{E}}_{x \sim \mu}\left[\mathrm{tr}(M^{\Psi_n} |\psi_x\rangle \langle \psi_x| \otimes |\psi_x\rangle \langle \psi_x|)\right] = \mathrm{tr}(M^{\Psi_n} \tilde{\rho}), \tag{4.3}$$

where we have defined $\tilde{\rho}$ as

$$\tilde{\rho} = \tilde{\mathbb{E}}_{x \sim \mu}\left[|\psi_x\rangle \langle \psi_x| \otimes |\psi_x\rangle \langle \psi_x|\right]. \tag{4.4}$$

We note that this is precisely the state obtained by applying a partial trace to all but $\tilde{O}(\sqrt{n})$ subsystems of the state ρ_x defined above.

Now, we need to calculate the expectation value of M^{Ψ_n}, the POVM element corresponding to the "yes" outcome of the protocol. Recall from Proposition 4.4 that our protocol is obtained by parallel repetition of the protocol of [HM13], where the number of repetitions is constant. (More precisely the number of repetitions is $O(\log 1/\epsilon)$ but we neglect this dependence since we take ϵ to be a constant and we allow the constant in the $\Omega(n)$ degree to depend on ϵ.) Thus, M^{Ψ_n} is a linear combination of tensor products of a constant number of terms, each of which implements a randomly chosen test on one of the registers of the witness state. The complementary POVM element $1 - M^{\Psi_n}$ consists of a linear combination of tensor products, where each product contains at least

one "no" outcome of a test. To show that the state $\tilde{\rho}$ passes M^{Ψ_n} with certainty, it suffices to show that the expectation value of any such term is 0. Below, we verify this for each test.

1. **Symmetry and Product tests** These tests consist of applying the swap test to various pairs of registers in the state. Since $\tilde{\rho}$ is fully symmetric under any permutation of the indices, we pass these tests with certainty, i.e. $\mathrm{Tr}[(M^{\text{"no", symmetry test}} \otimes M^{\text{rest}})\tilde{\rho}] = 0$.
2. **Uniformity test** Recall that in the uniformity test, Arthur chooses a matching $\mathcal{M}$ on $[n]$, and the measures each subsystem in an orthonormal basis containing

$$|\pm\rangle_{ij} \equiv \frac{1}{\sqrt{2}}(|i\rangle \pm |j\rangle)$$

for every $(i, j) \in \mathcal{M}$. The test fails if for some (i, j), outcomes of different subsystems are different. We claim this won't happen with $\tilde{\rho}$. Without loss of generality, let the first two subsystems have different outcomes. The probability for this to happen is given by

$$\begin{aligned}
&\mathbf{Pr}[\text{Uniformity test failure}] \\
&= \mathrm{Tr}[\tilde{\rho}\, |+\rangle_{ij} \langle +|_{ij} \otimes |-\rangle_{ij} \langle -|_{ij} \otimes M^{\text{rest}}] \\
&\propto \tilde{\mathbb{E}}_{x\sim\mu}[(x_i + x_j)^2 (x_i - x_j)^2 q(x)], \text{ for some polynomial } q(x) \\
&= \tilde{\mathbb{E}}_{x\sim\mu}[(x_i^2 + x_j^2 + 2x_i x_j)(x_i^2 + x_j^2 - 2x_i x_j) q(x)] \\
&= \tilde{\mathbb{E}}_{x\sim\mu}[(2 + 2x_i x_j)(2 - 2x_i x_j) q(x)] = \tilde{\mathbb{E}}_{x\sim\mu}[4(1 - x_i^2 x_j^2) q(x)] = 0.
\end{aligned}$$

In the above calculation, we used (2.6) repeatedly to simplify the terms. Also note that since there are $\tilde{O}(\sqrt{n})$ registers in the witness state, the degree of $q(x)$ is $\tilde{O}(\sqrt{n})$, which is less than the degree of the pseudoexpectation $\Omega(n)$.
3. **Satisfiability test:** In the satisfiability test, we choose a set of clauses to measure that have no variables in common with each other. Now, we perform the following procedure on the witness: first perform a measurement to project the witness into the subspace $\mathrm{Span}(\{|i\rangle : i \in C\})$ spanned by the variables occurring in a clause C.
If we end up in the subspace associated with $C = 2\text{-OUT-OF-}4(x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4})$, then we perform another projective measurement to check that the state is orthogonal to

$$|C\rangle = \frac{1}{2}(|i_1\rangle + |i_2\rangle + |i_3\rangle + |i_4\rangle).$$

Let $\Pi_C = |C\rangle\langle C|$. Let m be the number of copies of the witness, and suppose that the first stage of this test projects us onto clauses $C_1, \ldots, C_m$. Then the probability of the second stage passing is

$$\begin{aligned}
\mathbf{Pr}[\text{Success}] &= \mathrm{Tr}\left[(I - \Pi_{C_1}) \otimes (I - \Pi_{C_2}) \otimes \cdots \otimes (I - \Pi_{C_m}) \otimes M^{\text{rest}} \tilde{\rho}\right] \\
&\propto \tilde{\mathbb{E}}_{x\sim\mu}\left[(1 - \frac{1}{4}(\sum_{x\in C_1} x)^2) \ldots (1 - \frac{1}{4}(\sum_{x\in C_m} x)^2) \ldots\right].
\end{aligned}$$

Now, we know that the pseudo-solution μ has degree $\Omega(n)$ and satisfies all the 2-OUT-OF-4 constraints. In particular for every clause $\tilde{\mathbb{E}}_x[(\sum_{x\in C_1} x) q(x)] = 0$ for all polynomials $q(x)$ with degree $o(n)$. But in the expression above, we have a product

of $m = \tilde{O}(\sqrt{n})$ terms, each of degree 2, so every term containing a factor of $\sum_{x \in C} x$ will vanish under the pseudo-expectation. This leaves us with $\mathbf{Pr}[\text{Success}] = 1$ as desired.
Similarly, if we end up in the subspace associated with $C = \text{EQ}(x_{i_1}, x_{i_2})$. We do the same thing except now we choose

$$|C\rangle = \frac{1}{\sqrt{2}}(|i_1\rangle - |i_2\rangle).$$

The analysis is analogous to the above and we end up with $\mathbf{Pr}[\text{Success}] = 1$ as desired.

Note that it is crucial that even in the parallel-repeated protocol, the number of subsystems is $\tilde{O}(\sqrt{n})$. This means that all the tests in the protocol translate to polynomials of degree at most $\tilde{O}(\sqrt{n})$ under the pseudoexpectation. Since the pseudoexpectation is valid up to degree $\Omega(n)$, this means that the tests cannot tell that $\tilde{\rho}$ is not an honest witness. □

We can understand Proposition 4.7 as an explicit lower bound on the Doherty–Parrilo–Spedalieri [DPS03] hierarchy for h_{Sep}.

Corollary 4.8. *For any constant ϵ, there exists a family of measurements M_d acting on a bipartite Hilbert space with local dimension d, such that $h_{\text{Sep}}(M) \leq \epsilon$, but the kth-level of DPS estimates this value to be 1, i.e.,* $\text{DPS}_k(M) = 1$ *for* $k \leq o(\log d/\text{polylog}\log d)$.

Proof. We take M_d to be M^{Ψ} from the QMA(2) protocol, and $\tilde{\rho}$ from (4.4). The state $\tilde{\rho}$ arises as the reduced density matrix of a fully symmetric state ρ on $\Omega(n)$ registers and is thus a $\tilde{O}(\sqrt{n})$-extendible state. Moreover, since ρ is invariant under *all* permutations of indices (including those that exchange row and column indices), it is *a fortiori* invariant under partial transposes, and hence PPT. Thus, ρ lies within the set of states explored by DPS at level $k = \tilde{O}(\sqrt{n})$. The statement follows because $d = 2^{\tilde{O}(\sqrt{n})}$. □

In [BBH+12] it was shown that computing the $2 \to 4$ norm was a special case of computing h_{Sep} and that in turn there was an approximation-preserving reduction from h_{Sep} to the $2 \to 4$ norm. Examining that construction, we see that it is $O(1)$-degree and this lets us immediately obtain the following bound.

Corollary 4.9. *The SoS relaxation needs at least* $\Omega(\log(d)/\text{polylog}\log(d))$ *levels to approximate* $\|A\|_{2\to 4}$ *up to multiplicative error of* $C = O(1)$.

4.3. DPS lower bounds from other protocols. There are two other ETH-based hardness results that we can also make unconditional, but we will only sketch the proof here. Both of these results are obtained by an argument similar to the proof of Corollary 4.8, but with a different choice of Constraint Satisfaction Problem (CSP) in place of 2- OUT- OF- 4- SAT, and a correspondingly different reduction from 3XOR and choice of QMA(2) or QMA(k) protocol.

The first result is an SDP hardness result for $(1, 1 - \tilde{O}(1/n))$ approximations to $h_{\text{Sep}(n,n)}$, using a protocol of [LNN12].

Theorem 4.10. *Let* QMA(2)- HONEST *be the problem of maximizing the acceptance probability of the protocol of [LNN12] over honest strategies (to be described below). Then for every n, there exists an instance of the problem* QMA(2)- HONEST *on* poly(n)

variables with true optimum value $\leq 1 - \tilde{\Omega}(1/n)$. *Moreover, there exists a pseudosolution that achieves value 1 on this problem. As a consequence, we obtain a family of measurements* M_d *with* $h_{\mathrm{Sep}}(M) \leq 1 - \tilde{\Omega}(1/n)$, *but for which* $\mathrm{DPS}_k(M) = 1$ *for* $k \leq o(n)$.

Proof. The protocol of [LNN12] solves an NP-hard graph coloring problem in QMA(2) with completeness 1 and soundness $1-\tilde{\Omega}(1/n)$. Schematically, the proof of the hardness result is:

$$3\mathrm{XOR}(n) \Longrightarrow \textsc{Graph- 3- coloring}(n) \Longrightarrow \mathrm{QMA}(2)\text{- }\textsc{Honest} \Longrightarrow h_{\mathrm{Sep}(n,n)},$$

where $\textsc{Graph- 3- coloring}(n)$ is the problem of deciding whether a graph of n vertices is 3-colorable, and QMA(2)- HONEST is defined as before but with reference to the honest witnesses of the protocol of [LNN12]. To achieve hardness for h_{Sep}, we need to show that the reductions represented by the first two arrows of the diagram is pseudosolution-preserving, and that the last arrow is an approximate embedding reduction. The first arrow is a standard construction (see for instance proposition 2.27 of the textbook by Goldreich [Gol08]), quite similar to the gadget reduction for 2- OUT- OF- 4- SAT we considered earlier. It is straightforward to verify that it satisfies the hypothesis of proposition 3.16 with a constant value of κ; hence, it is pseudosolution preserving.

For the second arrow, we use a strategy similar to the proof of Proposition 4.7, arguing that a pseudosolution to $\textsc{Graph- 3- coloring}(n)$ can be turned into a dishonest quantum witness state $|\psi\rangle$, and that each test of the QMA(2) protocol evaluates a low-degree polynomial on the coefficients of $|\psi\rangle$, and thus passes with certainty. One difference from the previous case is that $\textsc{Graph- 3- coloring}(n)$ is not strictly speaking a Boolean problem. We remedy this by considering colorings that are *induced* through the reduction from an underlying assignment x to the variables 3XOR instance. For every vertex i and color c, let $g_{c,i}(x)$ be equal to 1 if vertex i is colored with c in this induced coloring, and 0 otherwise. Then, due to the locality of the gadgets in the reduction from 3XOR, the functions $g_{c,i}(x)$ are constant-degree polynomials in the variables x. Next, we need to verify that the connection between pseudosolutions and dishonest quantum witness states still holds in this protocol. Indeed, the honest witness state in this protocol for a given assignment x is $|\Psi_x\rangle\langle\Psi_x| = (|\psi_x\rangle\langle\psi_x|)^{\otimes 2}$, where

$$|\psi_x\rangle = \sum_{i,c} g_{c,i}(x)\,|i\rangle\,|c\rangle\,.$$

Here the index i runs over the vertices of the graph. Following our strategy in 4.7, we choose the following dishonest quantum witness state

$$\rho = \tilde{\mathbb{E}}_x[|\psi_x\rangle\langle\psi_x| \otimes |\psi_x\rangle\langle\psi_x|].$$

That this is indeed a valid quantum state follows from the properties of the pseudo-expectation operator, and from Lemma 3.15 (the functions $g_{c,i}(x)$ play the role of the polynomial map p in the lemma). This gives us a degree-$\Omega(n)$ value 1 pseudo-solution for QMA(2)- HONEST. Finally, the reduction in the last arrow is a $(1-\tilde{O}(1/n), 1-\eta)$-approximate embedding reduction for some constant η, by the soundness of the protocol. This yields a degree-$\Omega(n)$ lower bound for SoS approximations $h_{\mathrm{Sep}(n,n)}$ achieving approximation factor $(1, 1-\tilde{O}(1/n))$. □

The second result applies for multipartite separability. To obtain it, we replace the protocol of [HM13] with that [CD10], which is a QMA($O(\sqrt{n})$) protocol for 3-SAT with completness $1-\exp(-\Omega(\sqrt{n}))$ and soundness $1-\Omega(1)$, and which only performs Bell measurements (i.e. each party measures individually and then the outcomes are classically processed). We use this to prove hardness for the problem $h_{\mathrm{Sep}^{O(\sqrt{n})}(n)}(M)$, where $\mathrm{Sep}^k(n)$ means k-partite separable states, and M is restricted to be a bell measurement. The schematic diagram for this case is

$$3\mathrm{XOR}(n) \implies \textsc{Graph- 3- coloring}(n) \implies \mathrm{QMA}(2)\text{- }\textsc{Honest} \implies h_{\mathrm{Sep}^{O(\sqrt{n})}(n)}(M).$$

The arguments are very similar to those in the previous result; when we work out the parameters, we obtain that SoS needs at least $\Omega(n)$ levels to achieves a $(1-\exp(-\Omega(\sqrt{n})), 1-\eta)$ approximation to $h_{\mathrm{Sep}^{O(\sqrt{n})}(n)}(M)$ for general Bell measurements M, where η is an appropriately chosen constant.

Theorem 4.11. *For a sufficiently small constant η, there exists a family of measurements M_n over $O(\sqrt{n})$-partite states with local dimension n, such that $h_{\mathrm{Sep}^{O(\sqrt{n})}(n)}(M) \leq 1-\eta$, but $\mathrm{DPS}_k(M) \geq 1-\exp(-\Omega(\sqrt{n}))$ for all $k \leq o(n)$.*

5. SDP extended formulation Lower Bounds for h_{Sep}

In this section, we will leverage our SoS lower bounds to prove a lower bound on the size of *any* SDP relaxation approximating h_{Sep}, provided the relaxation is of a certain type called an extended formulation. To that end, we make use of a recent result of Lee, Raghavendra, and Steurer that relates extension complexity to SoS degree. We start by reviewing the techniques used by them.

5.1. SDP extended formulation lower bounds from LRS. In this section we illustrate the recent celebrated result of Lee, Raghavendra, and Steurer (LRS) [LRS15] lower-bounding the size of SDP extended formulations of boolean polynomial optimization problems in terms of the sum-of-squares degree. We will restate their main result (in a slightly more general form) in our current framework. Note that the LRS result plays a crucial role in extending our results on SoS hierarchies to more general SDP relaxations. Our contribution to this topic can be viewed more broadly as developing techniques to apply LRS to general optimization problems.

To that end, we first define the notion of SDP extended formulations as follows.

Definition 5.1. (*SDP*) A semidefinite program (SDP) A is an optimization problem, parametrized by $n \in \mathbb{N}$, with the following restrictions.

- The feasible set $\mathcal{P}_n^A$ is a spectrahedron contained in $L_{\geq}(\mathbb{R}^r)$, where $r = r(n)$ is called the size of this SDP and $L_{\geq}(V)$ denotes the set of positive-semidefinite matrices. By spectrahedron we mean simply a space of the form $W \cap L_{\geq}(\mathbb{R}^r)$ for W an affine subspace of $L(\mathbb{R}^r)$.
- Any instance Φ_n^A is an affine function from $L(\mathbb{R}^r)$ to $[0, 1]$.

Definition 5.2. (SDP extended formulation) For any optimization problem A, an SDP B is called a (c, s)-*approximate extended formulation* of A if there exists an *embedded* reduction $R_{A\Rightarrow B}$ that is $(s^B = c, s^A = s)$-*approximate*.

We note that the SDP extended formulation defined above is a more stringent concept than the conventional notion of an SDP *relaxation*, both because of the embedding property and because we require that the constraints do not depend on the objective function. The intuition behind this definition is that one can decide whether a given instance of problem A has value $\geq c$ (the "yes" case) or $< s$ (the "no" case) by solving the SDP corresponding to (c, s)-approximate extended formulation of the problem A; if the objective value of the SDP is at least c, one concludes that one is in the "yes" case, and otherwise, that one is in in the "no" case.

It is not hard to see that any SDP extended formulation for B is also a SDP extended formulation for A if there is an embedded reduction from A to B with matching approximation parameters. Precisely,

Proposition 5.3. *Let a SDP C be a (s^C, s^B)-approximate SDP extended formulation of an optimization problem B. If there is an embedded reduction $R_{A\Rightarrow B}$ that is (s^B, s^A)-approximate, then C is a (s^C, s^A)-approximate SDP extended formulation of the optimization problem A.*

Proof. This claim follows by definition and the transitivity of approximating and embedding properties of reductions in Proposition 3.13. □

The above definitions allow us to translate lower bounds on SDPs for one problem to another. To obtain such lower bounds in the first place, we use a technique developed by [LRS15], which bounds the positive semidefinite rank of a particular "pattern matrix." In particular, suppose we would like to give an SDP lower bound on (c, s)-approximations for an optimization problem. The first observation that LRS make is that to achieve this, it suffices to lower bound the psd rank of the matrix $M^n_{(c,s)}$, whose rows are indexed by instances Φ whose true optimum value is $\leq s$, and whose columns are indexed by feasible points $x \in \{0, 1\}^n$. The value of an entry of this matrix is given by $M^n_{(c,s)}(\Phi, x) = c - \Phi(x)$. Note that all entries of this matrix are nonnegative by construction, so the psd rank is well defined. The second key observation of LRS is that there is a relation between SoS degree of the function $c - \Phi(x)$, and the psd rank of a *different* matrix

$$M^n_\Phi : [n]^m \times \{0, 1\}^n \mapsto \mathbb{R}_{\geq 0},\ M^n_\Phi(S, x) = c - \Phi(x_S).$$

Here, the rows are indexed by subsets S of size m, and the notation x_S means the values of x on the coordinates in the subset S. The key technical lemma of LRS is the following:

Lemma 5.4. (Theorem 3.8 of [LRS15]) *Suppose Φ is an instance of an optimization problem over m variables, and* $\deg_{\mathrm{SoS}}(c - \Phi(x)) \geq d$. *Then for $n \geq m^{d/4}$,* $\mathrm{rk}_{\mathrm{psd}}(M^n_\Phi) \geq \Omega(m^{d^2/8})$.

To relate this to the original problem, we need to show that M^n_Φ is a *submatrix* of $M^n_{(c,s)}$.

While this approach is so far the most successful route to general SDP lower bounds, it has two limitations. First, the requirement that $n \geq m^{d/4}$ implies that the lower bound on psd rank (and hence SDP size) obtained will never be better than *quasi-polynomial* in n. This requirement seems essential for the random restriction analysis which is central to the proof of LRS. This means that the bounds obtained via this method can be much looser than the SoS lower bounds they are based on. A second limitation appears when we try to use the technique for settings *other* than CSPs. Essentially, the problem is that we need to interpret an instance of the problem on m variables as an instance on $n \gg m$

variables, in order for the matrix M^n_Φ to be a submatrix of $M^n_{(c,s)}$. This is straightforward in the case of CSPs, but not for other problems. For instance, the problems we will consider in this work arise from particular quantum proof protocols; these protocols involve states that are superpositions over all of the variables in the problem, and as a result break down when only a small number of variables enter into the objective function. As a result of these limitations, we will only be able to obtain SDP lower bounds in *some* of the cases where we have SoS lower bounds, and even in those cases, our parameters will be worse than those of the SoS results. We consider it a major open problem to improve on these techniques and prove tighter SDP lower bounds.

5.2. Applying LRS to h_{Sep}. To apply the LRS techniques to h_{Sep}, we need to re-examine each reduction in our chain of reductions to ensure the embedding property. First, we need to investigate in more detail the 3-coloring proof system of [LNN12], which consists of the following steps:

1. The verifier receives a product state $|\psi_A\rangle \otimes |\psi_B\rangle$ from the two provers. Each prover's state consists of one register of $\log(n)$ qubits, holding an index from 1 to n, and a second register consisting of a single qutrit, whose three states correspond to the three possible colors in the graph.
2. The verifier performs one of the following tests:
   - **Uniformity test** For each proof state, the verifier performs a quantum Fourier transform on the color register, and measures it. If he obtains 0, then he performs an inverse Fourier transform on the index register, and measures it. He accepts if he measures 0 and rejects otherwise.
   - **Satisfiability test** The verifier measures both proof states in the computational basis, obtaining two tuples (i, c_i) and (j, c_j) of indices and colors. If there is no edge between i and j in the graph, then the verifier accepts. If there is an edge, then the verifier accepts if $c_i \neq c_j$ and rejects otherwise.

For this protocol, the honest witness states are those of the form

$$|\Psi\rangle = \left(\frac{1}{\sqrt{n}} \sum_i |i\rangle \otimes |c_i\rangle\right)^{\otimes 2}.$$

When we try to apply LRS directly to the problem QMA(2)- HONEST for this protocol, we run into several obstacles:

1. LRS requires that the feasible set of the optimization problem be the entire Boolean cube $\{0, 1\}^n$. This means that we cannot take the problem QMA(2)- HONEST to be an optimization over all colorings, but rather we must restrict ourselves to colorings *induced* by Boolean assignments to the variables of an underlying 3XOR instance, as we did in the proof of Theorem 4.10.
2. The embedding property of LRS means that the SDP feasible point corresponding to an assignment x must be *independent* of the objective function, i.e. the choice of graph. However, in the construction given in the proof of Theorem 4.10, the induced coloring depends on the graph.

To address these issues, we make some tweaks to the protocol. First, for every input size n, we choose a *universal* graph G_n, which is induced by a 3XOR instance with a *complete* constraint graph. For every assignment x to the 3XOR variables, we let the

induced coloring c be the coloring induced on the universal graph G_n. It is not hard to see that, in the standard gadget reduction, the graph obtained will be a subgraph of this universal graph G_n, and the induced coloring will match the universal induced coloring.

Now, using this modified protocol, we can prove our main result. First, we state the soundness property of the protocol in a form that will be useful to us.

Lemma 5.5. (Soundness analysis of [LNN12]) *There exist a constant $\eta < \frac{1}{2}$ such that if Φ is a* 3XOR *instance on n variables with $O(n)$ clauses and value at most $1-\eta$,*[5] *then the LNN protocol accepts with probability at most $1-\Omega(1/(n\text{polylog}n))$. Moreover, for any unsatisfiable Φ on n variables, then the acceptance probability of LNN is at most $1-\Omega(1/n^2)$.*

Theorem 5.6. *Any SDP extended formulation achieving a $(1-\epsilon(d), 1-\delta(d))$-approximation to $h_{\text{Sep}(d,d)}$ where $\delta(d) = O(1/d^2)$ and $\epsilon(d) < \delta(d)$ has size at least $d^{\log d/\text{poly}\log\log d}$.*

Proof. The proof follows the strategy outlined in Sect. 5.1. First we will rule out efficient SDP extended formulations to QMA(2)- HONEST, and then show that this implies the nonexistence of SDP extended formulations for h_{Sep} as well.

To do this, we use the pattern matrix technique of LRS. Let $\Phi_m(x) : \{0,1\}^m \to \mathbb{R}_{\geq 0}$ be the objective function of QMA(2)- HONEST on instances of size m, induced by the hard 3XOR instance of Grigoriev. Then we know that $\max_x \Phi_m(x) \leq 1-1/(n\text{polylog}n)$, and $\deg_{\text{SoS}}(1-\Phi_m(c)) \geq \Omega(m)$, i.e. there is a degree-$\Omega(m)$ pseudodistribution under which $\tilde{\mathbb{E}}[(1-\Phi(c))] = 0$. Now, by Lemma 5.4, this implies that the matrix $M^n_{\Phi_m}(S,x) = 1-\epsilon(n)-\Phi_m(x_S)$ has PSD rank at least $n^{\Omega(m)}$ for $n = m^{\Omega(m)}$. We would like this to be a submatrix of $M_{c,s}(\Phi_n, x) = c - \Phi_n(x)$ for some choice of c, s, where the instances Φ_n are now over n variables. However, this would require that we be able to "simulate" the action of the protocol on m variables using larger instances on n variables, and achieve exactly the same objective value. Since the LNN protocol involves sampling variables from witness states in uniform superposition, it is not obvious how to do this—if we run the protocol over n variables, then most of the samples will lie outside any particular subset of size m of the variables.

In order to avoid this obstacle, we define a new function $f'(x) : \{0,1\}^m \to \mathbb{R}_{\geq 0}$, which is equal to the success probability of the LNN protocol when the witness state is $|\Psi_{x'}\rangle$ where $x' \in \{0,1\}^n$ agrees with x on the first m coordinates, and when the verifier is given the Grigoriev 3XOR instance applied only to the first m coordinates of x'. The soundness properties in Lemma 5.5 imply that $\max_x f'(x) \leq 1-\Omega(1/n^2) := 1-\delta(n)$. At the same time, the Grigoriev pseudosolution also yields a pseudosolution for which $\tilde{\mathbb{E}}_x[f'(x)] = 1$. Hence, $\deg_{\text{SoS}}(1-f'(x)) \geq \Omega(m)$, and thus by Lemma 5.4, the matrix

$$M^{f'}_n(S,x) = 1-\epsilon(n) - f'(x_S)$$

has PSD rank at least $n^{\Omega(m)}$. Now, *this* matrix is indeed an exact submatrix of $M_{c,s}(\Phi_n, x)$ as defined above for $c = 1-\epsilon(n)$, $s = 1-\delta(n)$. So any SDP extended formulation to QMA(2)- HONEST has size at least $n^{\Omega(m)} = n^{\log n/\text{poly}\log\log n}$.

Now, note that the trivial reduction from QMA(2)- HONEST to h_{Sep} is an *embedding* reduction. Moreover, we claim that it is $(1-\delta(n), 1-\delta(n))$-approximate. To see this, note that whenever QMA(2)- HONEST $\leq 1-\delta(n)$, the underlying 3XOR instance must

[5] In [LNN12], these properties are obtained by applying Dinur's PCP theorem.

be infeasible, and thus $h_{\text{Sep}} \leq 1-\delta(n)$ by Lemma 5.5. Finally, it remains to translate our result into a bound in terms of the dimension of the state d. Recall that the dimension d of the h_{Sep} instance is polynomially related to the number of variables n, since the witness states consist of $O(\log n)$ qubits. Hence, by the previous result for QMA(2)- HONEST and Proposition 5.3, we conclude that for an appropriate $\delta'(d) = O(1/d^2)$, any $(1-\epsilon'(d), 1-\delta'(d))$-approximate SDP extended formulation to $h_{\text{Sep}(d,d)}$ with $\epsilon'(d) < \delta'(d)$ must have size at least $d^{\log d/\text{polylog}\log d}$. □

Using the equivalence between $2 \to 4$ norm and h_{Sep}, we can likewise obtain an SDP hardness result for the $2 \to 4$ norm.

Corollary 5.7. *Any SDP extended formulation to $\|A\|_{2\to 4}$ for d-dimensional tensors achieving a multiplicative error of $C = 1/O(d^2)$ must have size at least $\Omega(d^{\log(d)/\text{polylog}\log(d)})$.*

5.3. The no-disentangler conjecture. One application of our result is to prove a version of the Approximate Disentangler Conjecture for a particular range of parameters. This conjecture was originally formulated by Watrous and first published in [ABD+08]. Previously the only evidence in favor of this conjecture was based on complexity assumptions (e.g. the ETH) and even those results did not rule out the possibility of disentangling maps that were hard to compute.

Definition 5.8. Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, and denote the space of density matrices on $\mathcal{H}$ by $D(\mathcal{H})$ (likewise $D(\mathcal{K})$). A linear CPTP map $\Lambda : D(\mathcal{H}) \to D(\mathcal{K} \otimes \mathcal{K})$ is an (ϵ, δ)-*approximate disentangler* if

- For every $\rho \in D(\mathcal{H})$, $\Lambda(\rho)$ is ϵ-close in trace distance to a separable state in $\text{Sep}(\mathcal{K} \otimes \mathcal{K})$.
- For every separable state $\sigma \in \text{Sep}(\mathcal{K} \otimes \mathcal{K})$, there exists a $\rho \in D(\mathcal{H})$ such that $\Lambda(\rho)$ is δ-close in trace distance to σ.

Our result is the following:

Theorem 5.9. *Let $d = \dim(\mathcal{K})$, and suppose that $\Lambda : D(\mathcal{H}) \to D(\mathcal{K} \otimes \mathcal{K})$ is an (ϵ, δ)-approximate disentangler with $\epsilon + \delta < 1/\text{poly}(d)$. Then*

$$\dim(\mathcal{H}) \geq \Omega(d^{\log(d)/\text{polylog}\log(d)}).$$

This is considerably weaker than Watrous's original formulation, which had $\epsilon+\delta < O(1)$ and $\dim(\mathcal{H}) \geq \exp(d)$. Conditional on ETH, the result was known with $\epsilon+\delta < O(1)$ and $\dim(\mathcal{H}) \geq d^{\log(d)/\text{polylog}\log(d)}$; we are unable to match this due to technical limitations of the LRS result, which are described in more detail in the previous subsection.

Proof. We show that a disentangler can be used as an SDP extended formulation to h_{Sep}, thus allowing us to apply Theorem 5.6. Throughout the proof, let $d \equiv \dim(\mathcal{K})$. First, let us consider the $\delta = 0$ case. We define the optimization problem

$$h_\Lambda(M) \equiv \max_{\rho \in D(\mathcal{H})} \text{Tr}[M\Lambda(\rho)].$$

Note that $h_\Lambda(M)$ is a semidefinite program with size $\dim(\mathcal{H})$. Moreover, when $\delta = 0$, we claim that there exists an embedding reduction R_Λ from h_{Sep} to $h_\Lambda(M)$, that

achieves a $(s+\epsilon, s)$ approximation. This reduction simply maps an instance M of h_{Sep} to the instance $h_\Lambda(M)$ given by the same measurement operator M. The embedding property follows from the definitions: for every separable σ, there exists a ρ such that $\Lambda(\rho) = \sigma$, and so $\text{Tr}[M\Lambda(\rho)] = \text{Tr}[M\sigma]$. Similarly, the soundness of the reduction also follows from the definition: for every ρ, $\Lambda(\rho)$ is ϵ-close to some separable σ. This means that $\max_\rho \text{Tr}[M\Lambda(\rho)] \leq \max_{\sigma\in\text{Sep}} \text{Tr}[M\sigma] + \epsilon$. Consequently, $h_\Lambda(M)$ is a SDP extended formulation of h_{Sep}. So, by applying theorem 5.6, we conclude that $\dim(\mathcal{H}) \geq d^{\Omega(log(d)/\text{polylog}\log(d))}$.

Now, let us consider the general case, where $\delta > 0$. In this case, we cannot directly apply the preceding argument, since there is no embedding from h_{Sep} to $h_{Disentangled}$. We will fix this by using the following gadget: Let $\mathcal{B}_\delta$ be the set of states in $D(\mathcal{K}\otimes\mathcal{K})$ of trace norm less than or equal to δ. Then, given an (ϵ, δ)-disentangler Λ, we define a new map $\tilde{\Lambda} : D(\mathcal{H}) \oplus \mathcal{B}_\delta \to D(\mathcal{K}\otimes\mathcal{K})$ by

$$\tilde{\Lambda}(\rho \oplus \sigma) \equiv \Lambda(\rho) + \sigma.$$

We claim that for every separable $\tau \in \text{Sep}(\mathcal{K}, \mathcal{K})$, there exists a preimage $\rho \in D(\mathcal{H}), \sigma \in \mathcal{B}_\delta$ with $\tilde{\Lambda}(\rho\oplus\sigma) = \tau$. Indeed, since Λ is an (ϵ, δ)-disentangler, we know that τ had an *approximate* preimage ρ satisfying $\Lambda(\rho) = \tau + \sigma$ for $\|\sigma\|_1 \leq \delta$. From our definition of $\tilde{\Lambda}$ it follows that $\tilde{\Lambda}(\rho\oplus\sigma) = \tau$ as desired. We also claim that for every $\rho \in D(\mathcal{H}), \sigma \in \mathcal{B}_\delta$, $\tilde{\Lambda}(\rho\oplus\sigma)$ is within $\epsilon+\delta$ in trace distance of some separable state. To see this, note that $\Lambda(\rho)$ is within ϵ distance of some separable state τ, and since $\|\sigma\|_1 \leq \delta$, adding σ can increase the distance to τ by at most σ.

These two claims tell us that $\tilde{\Lambda}$ is "almost" an $(\epsilon+\delta, 0)$-approximate disentangler: the only catch is that it is not a CPTP map acting on quantum states. Nevertheless, we can still use the same argument as the $\delta = 0$ case. We define the optimization problem $h_{\tilde{\Lambda}}(M)$ by the SDP

$$\begin{aligned} \max_{\rho,\sigma^+,\sigma^-} \quad & \text{Tr}[M\tilde{\Lambda}(\rho\oplus\sigma)] \\ \text{such that} \quad & \text{Tr}[\rho] = 1 \\ & \text{Tr}[\sigma^+ + \sigma^-] \leq \delta \\ & \rho, \sigma^+, \sigma^- \succeq 0. \end{aligned}$$

This SDP implements the constraint $\|\sigma\|_1 \leq \delta$. As before, consider the reduction from h_{Sep} to $h_{\tilde{\Lambda}}$ that maps the instance M to the instance corresponding to the same measurement operator M. We claim that this reduction is an embedding and is $(s, s+\epsilon+\delta)$ approximate for any s; these claims are proved by a similar argument to the $\delta = 0$ case. Thus, we have shown that $h_{\tilde{\Lambda}}$ is a $(s, s+\epsilon+\delta)$-approximate SDP extended formulation of h_{Sep}. So once again, applying Theorem 5.6 tells us that the dimension of the SDP $h_{\tilde{\Lambda}}$ must be at least $\Omega(d^{\log d/\text{polylog}\log d})$. Now, the dimension of $h_{\tilde{\Lambda}}$ is equal to $\dim(\mathcal{H}) + d$, so all together we get

$$\dim(\mathcal{H}) \geq \Omega(d^{\log d/\text{polylog}\log d}) - d \geq \Omega(d^{\log d/\text{polylog}\log d}).$$

□

6. Lower Bounds for Entangled Games

In this section, we show lower bounds on SDP extended formulations for the entangled value of quantum games. First, we review some basic notions.

Definition 6.1. A *nonlocal game* $G = (Q, A, \pi, V)$ is a game played between a referee, or "verifier," and two players, or "provers." In one round, the verifier chooses two random questions $q_1, q_2 \in Q$ according to the joint probability distribution $\pi(q_1, q_2)$, and sends q_1 to player 1 and q_2 to player 2. Each player returns an answer in the set A. The verifier then *accepts* the provers' answers with probability given by $V(q_1, q_2, a_1, a_2)$. The *winning probability* of a strategy is the probability that the verifier accepts when the players play according to the strategy.

Strategies can be either classical or entangled.

Definition 6.2. A (deterministic) *classical strategy* for a nonlocal game consists of functions $f_1, f_2 : Q \to A$, with $f_1(q)$ being the answer that player 1 gives to question q, and likewise for $f_2(q)$. The *classical value* $\omega_{classical}(G)$ of a game G is the maximum winning probability of a classical strategy for G.

Equivalently, we could have allowed the two players to share classical random bits. However, by a simple convexity argument one can show that the classical value of a game is always achieved by a deterministic strategy.

Definition 6.3. A *quantum strategy* for a nonlocal game consists of:

- Hilbert spaces H_1, H_2 and a joint state $|\psi\rangle \in H_1 \otimes H_2$,
- for every question $q \in Q$, a POVM $\{A_q^a \otimes I_2\}_{a \in A}$ acting nontrivially on only player 1's Hilbert space, and
- for every question $q \in Q$, a POVM $\{I_1 \otimes B_q^a\}_{a \in A}$ acting nontrivially on only player 2's Hilbert space.

To play the game, each player measures their shared state $|\psi\rangle$ using the POVM associated with the question received, and returns the POVM outcome as the answer. The *entangled value* $\omega_{\text{entangled}}(G)$ of a game G is the maximum winning probability of an entangled strategy for G.

In this section, we show a lower bound on the size of an SDP to compute the entangled value of a 2-player entangled game, to within inverse polynomial accuracy. We show both a bound on the size of general SDP extended formulations, as well as an explicit integrality gap for the non-commuting SoS hierarchy. We do this by embedding 3XOR into a quantum entangled game, using a result of Ito, Kobayashi, and Matsumoto [IKM09].

$$\text{3XOR} \underset{R_1}{\Longrightarrow} \omega_{\text{HONEST CLASSICAL}} \underset{R_2}{\Longrightarrow} \omega_{\text{entangled}}$$

The intermediate problem is

- The problem $\omega_{\text{HONEST CLASSICAL}}(G)$ is a boolean polynomial optimization problem. Each instance is parametrized by a 2-player game G of the form considered by [IKM09]. The objective function $f(x)$ in the optimization evaluates the winning probability in G of a classical strategy parametrized by a boolean string x.

Before we explain these reductions in more detail, we first review the result of [IKM09] that we will use.

Lemma 6.4. (Lemma 8 of [IKM09]) *Let Φ be a 3-CSP over n variables with m clauses.*[6] *Then there exists a 2-player quantum game G_Φ such that for some constant $\gamma > 0$,*

$$\text{MAX-SAT}(\Phi) \leq \omega_{classical}(G_\Phi) \leq 1 - \frac{1 - \text{MAX-SAT}(\Phi)}{3}$$
$$\omega_{entangled}(G_\Phi) \leq 1 - \frac{\gamma(1 - \text{MAX-SAT}(\Phi))^2}{m^2}.$$

The game G_Φ is constructed starting from Φ, using the technique of *oracularization with a dummy question*.

Definition 6.5. Let Φ be a 3-CSP. Then the oracularization of Φ is a 2-player entangled game G_Φ. In this game two random clauses are sampled from Φ, say acting on bits (i_1, i_2, i_3) and (i'_1, i'_2, i'_3), which we assume are randomly ordered. Then one player receives (i_1, i_2, i_3) and the other player receives with equal probability either (i_j, i'_1) or (i'_1, i_j) with j drawn randomly from $\{1, 2, 3\}$. The players answer with 3 and 2 bits respectively. The verifier then accepts if both of the following two checks pass:

1. **Simulation check** The verifier checks that the answers from player 1 satisfy the clause associated with variables i_1, i_2, i_3.
2. **Consistency check** For the variable i_j, the verifier checks that both players' answers for this variable agree.

In our application of this result, we will take the 3-CSP to be 3XOR. We say that the players are playing *honestly according to assignment* x if the players responds with the answers $(x_{i_1}, x_{i_2}, x_{i_3})$ and either $(x_{i_j}, x_{i'})$ or $(x_{i'}, x_{i_j})$ as appropriate. Thus, the consistency check will always pass and the simulation check will pass with probability $f_\Phi(x)$, which is defined to equal the fraction of clauses in Φ satisfied by x.

Definition 6.6. The problem $\omega_{\text{HONEST CLASSICAL}}(G_\Phi)$ is the optimization problem of maximizing $f_\Phi(x)$ over $x \in \{\pm 1\}^n$.

Lemma 6.7. *Let Φ be a* 3XOR *instance produced by Proposition 3.18. Then there exists a degree-$\Omega(n)$ value-$(1, \frac{1}{2} + \epsilon)$ integrality gap for $\omega_{\text{HONEST CLASSICAL}}(G_\Phi)$ for all ϵ.*

Proof. Essentially, this follows directly from the fact that $f_\Phi(x)$ counts the fraction of clauses satisfied by x, since the other tests in the game all pass with probability 1 for honest strategies.

In more detail, the function $f_\Phi(x)$ is a polynomial function of the variables x. Each term in the polynomial corresponds to a possible check that the verifier performs. We show that each term has pseudoexpectation 1 under the pseudoexpectation operator $\tilde{\mathbb{E}}$ produced by proposition 3.18. Recall that this pseudoexpectation operator has degree $\Omega(n)$.

- **Simulation test** In this test, we verify that the answers of each prover satisfy the clause they were asked. In other words, for a 3XOR clause $x_i x_j x_k = b$, we want to verify that the player's answers multiply together to b. For every clause $b = x_i x_j x_k$, we have a term

$$\frac{1}{2} + \frac{1}{2} b x_i x_j x_k,$$

[6] This is called a "nonadaptive 3-query PCP" in their language.

in the polynomial $f_\Phi(x)$. We compute the pseudoexpectation of this term:

$$\begin{aligned}\tilde{\mathbb{E}}'[V_{b,A}] &= \frac{1}{2} + \frac{1}{2} b \tilde{\mathbb{E}}[x_i x_j x_k] \\ &= 1.\end{aligned}$$

- **Consistency test** In this test, we check that the two players give the same answer when asked about the same bit. This test is automatically satisfied for *any* input to $f_\Phi(x)$, since any honest strategy is consistent. Thus, it is also satisfied by the pseudoexpectation $\tilde{\mathbb{E}}$.

Thus, we have shown that there exists a degree-$\Omega(n)$ pseudoexpectation $\tilde{\mathbb{E}}$ such that $\tilde{\mathbb{E}}[f_\Phi(x)] = 1$. However, notice that $f_\Phi(x) = \alpha\Phi(x) + \beta$, where α is the probability of doing a simulation test and β the probability of doing a consistency test. Thus, since MAX-3XOR$(\Phi) \leq \frac{1}{2}+\delta$, we deduce that $\max_x f_\Phi(x) \leq \alpha(\frac{1}{2}+\delta)+\beta = 1-\alpha(\frac{1}{2}-\delta) \leq 1-c$ for the appropriate constant c. Thus, $\tilde{\mathbb{E}}$ gives us the desired degree $\Omega(n)$, value-$(1-\epsilon, 1-c)$ integrality gap for all $\epsilon > 0$. □

6.1. General SDPs. First, we show an SDP lower bound for the optimization over honest strategies.

Lemma 6.8. *Suppose S_n is a sequence of SDP extended formulations to the problem $\omega_{\text{HONEST CLASSICAL}}(G)$ of size r_n, achieving an $(c = 1-\epsilon(n), s = 1-\delta)$-approximation, where $\delta < \frac{1}{2}$ and $\epsilon(n) < \delta$. Then $r_n \geq \Omega\left(n^{\log n/\text{poly}\log\log n}\right)$.*

Proof. The proof of this theorem is similar to that of Theorem 5.6, and also relies on the result of LRS. A notable simplification that occurs in the games setting is that it is *easier* to embed an instance of a problem into an instance of the same problem with more variables.

Let $f_\Phi(x)$ be the objective function of $\omega_{\text{HONEST CLASSICAL}}(G_\Phi)$ on an instance of G given by a Grigoriev 3XOR instance Φ on m variables. By Lemma 6.7, we know that $\deg_{\text{SoS}}(1-\delta-f_\Phi(x)) \geq \Omega(m)$. Thus, by the LRS theorem, for $n = m^{\Omega(m)}$, the matrix $M_n^f(S,x) = 1-\delta-f_\Phi(x_S)$ has PSD rank at least $n^{\Omega(m)}$. This matrix is a submatrix of $M_{c,s}(\mathcal{J},x) = 1-\delta-\langle\mathcal{J},x\rangle$ which is the pattern matrix of the SDP extended formulation. So the SDP extended formulation has size at least $n^{\Omega(m)} = n^{\log n/\text{poly}\log\log n}$. □

For a general 2-player game G_N of size N (i.e. for which the distribution π over questions and predicate V can be specified in N bits), define the optimization problem $\omega_{\text{entangled}}(G_N)$ to be the optimization of the winning probability of game G_N over all two-player entangled strategies.

Theorem 6.9. *Suppose S_N is a sequence of SDP extended formulations to the problem $\omega_{entangled}(G_N)$ of size r_N, achieving an $(c = 1-\epsilon(N), s = 1-\delta(N))$-approximation, where $\delta(n) = O(1/N)$ and $\epsilon(N) < \delta(N)$. Then $r_N \geq \Omega\left(N^{\log N/\text{poly}\log\log N}\right)$.*

Proof. As before, let G_Φ be the oracularized game associated with the 3XOR instance Φ on n variables, and $f_\Phi(x)$ be the winning probability of an honest strategy played according to assignment x. From the definitions, it follows that there is an embedding reduction from the problem $\omega_{\text{HONEST CLASSICAL}}(G)$ to $\omega_{\text{entangled}}(G_N)$ where $N = \Theta(n^2)$.

Moreover, by Lemma 6.4, this reduction is $(1 - c/n^2, 1 - c')$-approximate for constants c, c'. Thus, any SDP extended formulation of size r_N for $\omega_{\text{entangled}}(G_N)$ that achieves a $(1 - \epsilon(N), 1 - c/N)$ approximation implies an SDP extended formulation of size r_N for $\omega_{\text{HONEST CLASSICAL}}(G)$ that achieves a $(1 - \epsilon(N), 1 - c')$ approximation. Now, by Lemma 6.8, any such SDP extended formulation must have size at least $r_N \geq \Omega\left(N^{\log N/\log\log N}\right)$. □

6.2. An explicit lower bound for ncSoS. In the previous section we gave a lower bound on the size of SDP extended formulations for the problem of computing the entangled game value. We will now present an explicit lower bound for a family of SDP extended formulations called the *non-commuting sum of squares* (ncSoS) hierarchy, also referred to as the NPA hierarchy [NPA08,DLTW08]. Recall that in the sum-of-squares hierarchy, one optimizes a polynomial function $f(x)$ by optimizing $\tilde{\mathbb{E}}_{x\sim\mu} f(x)$ over pseudodistributions μ that obey certain constraints. Likewise, in the ncSoS hierarchy, the winning probability ω is viewed as a polynomial in non-commuting variables (corresponding to the quantum operators in the provers' strategy), and the game value is found by optimizing the *non-commuting pseudoexpectation* of this nc polynomial. A non-commuting pseudoexpectation satisfies conditions similar to an ordinary pseudoexpectation operator:

Definition 6.10. An degree-d ncSoS pseudoexpectation is a linear map $\tilde{\mathbb{E}}[\cdot]$ that maps nc polynomials in the provers' measurement operators $\{A_q^a\}, \{B_q^a\}$ to real numbers. This map satisfies the following properties:

- Normalization: $\tilde{\mathbb{E}}[I] = 1$.
- Positivity: for all polynomials $p(A, B)$ with degree at most $d/2$, $\tilde{\mathbb{E}}[p^\dagger p] \geq 0$.
- Commutation: for any operators A, B acting on different provers, $\tilde{\mathbb{E}}[q_1(x)(AB - BA)q_2(x)] = 0$ for all polynomials $q_1(x), q_2(x)$ with $\deg q_1 + \deg q_2 \leq d - 2$.

In the following theorem, we show that when the degree d is small enough, we can construct a non-commuting pseudoexpectation according to which every test in the game is satisfied with probability 1, even though the game value is less than $1 - 1/\text{poly}(n)$. We do this in a quite generic way, by constructing a commuting pseudoexpectation for the CSP instance underlying the game, and "lifting" this pseudoexpectation to a non-commuting pseudoexpectation for the game.

Theorem 6.11. *For every n there exists a two-player entangled game G with $O(n)$ questions and three-bit answers, such that $\omega_{entangled}(G) \leq 1 - c/n^2$ for some constant c, but there exists a pseudoexpectation of degree $Omega(n)$ according to which the game value is 1.*

Proof. Start with a 3XOR instance with maximum satisfiable fraction $1/2 + \epsilon$. Then Lemma 6.4 gives us the first part of the conclusion. For the second part, we explicitly construct the pseudodistribution using the Grigoriev instance. Let the two players be denoted A and B. Their strategies are given by POVMs $\{A_{i_1 i_2 i_3}^{a_1 a_2 a_3}\}$, $\{B_{j_1 j_2}^{b_1 b_2}\}$, where $i_1, \ldots, i_3, j_1, \ldots, j_3$ are indices of variables in the 3XOR instance. To specify a pseudodistribution, we need to assign values to every pseudoexpectation of words built out of these variables. We do so as follows: first, we impose the condition that the A and B operators are mutually commuting, and moreover that $A_{i_1 i_2 i_3}^{a_1 a_2 a_3} = C_{i_1}^{a_1} C_{i_2}^{a_2} C_{i_3}^{a_3}$, $B_{j_1 j_2}^{b_1 b_2} = C_{j_1}^{b_1} C_{j_2}^{b_2}$ where the operators $\{C_i^0, C_i^1\}$ form a projective measurement for every index i.

For convenience, we will henceforth work with the observables $C_i \equiv C_i^0 - C_i^1$; these square to identity. Now, let $\tilde{\mathbb{E}}$ be the Grigoriev pseudoexpectation operator for the 3XOR instance. We define an ncSoS pseudoexpectation $\tilde{\mathbb{E}}'$ as follows:

$$\tilde{\mathbb{E}}'[C_{i_1} \dots C_{i_k}] \equiv \tilde{\mathbb{E}}[x_{i_1} \dots x_{i_k}].$$

By construction, this pseudoexpectation satisfies all the ncSoS constraints. It is defined up to degree $\Omega(n)$. We now need to check that it achieves a game value of 1. The game consists of two kinds of checks: simulation and consistency.

- **Simulation test** In this test, we verify that the answers of each prover satisfy the clause they were asked. In other words, for a 3XOR clause $x_i x_j x_k = b$, we want to verify that player A's answers multiply together to b. For every clause $b = x_i x_j x_k$, we have a term

$$V_{b,A} = \left(\sum_{a_i, a_j, a_k} a_i a_j a_k A_{x_i x_j x_k}^{a_i a_j a_k} \right) \otimes I_B,$$

  in the game value, and an analogous term for player B. We compute the pseudoexpectation of this term:

$$\begin{aligned}
\tilde{\mathbb{E}}'[V_{b,A}] &= \sum_{a_i, a_j, a_k} (a_i a_j a_k b) \tilde{\mathbb{E}}'[A_{x_i x_j x_k}^{a_i a_j a_k} \otimes I_B] \\
&= \sum_{a_i a_j a_k} a_i a_j a_k b \tilde{\mathbb{E}}'[C_{x_i}^{a_i} C_{x_j}^{a_j} C_{x_k}^{a_k} \otimes I_B] \\
&= b \tilde{\mathbb{E}}'[C_{x_i} C_{x_j} C_{x_k} \otimes I_B] \\
&= b \tilde{\mathbb{E}}[x_i x_j x_k] \\
&= 1.
\end{aligned}$$

- **Consistency test** In this test, we check that players A and B give the same answer if asked about the same bit.

$$\begin{aligned}
V_{i,A,B} &= \mathbb{E}_{j,k,p} \sum_{a_1 a_2 a_3} \sum_{b_1 b_2} (a_1 b_1) A_{x_i x_j x_k}^{a_1 a_2 a_3} \otimes B_{x_i x_p}^{b_1 b_2} \\
\tilde{\mathbb{E}}'[V_{i,A,B}] &= \mathbb{E}_{j,k,p} \sum_{a_1 a_2 a_3} \sum_{b_1 b_2} (a_1 b_1) \tilde{\mathbb{E}}'[A_{x_i x_j x_k}^{a_1 a_2 a_3} \otimes B_{x_i x_p}^{b_1 b_2}] \\
&= \mathbb{E}_{j,k,p} \sum_{a_1 a_2 a_3} \sum_{b_1 b_2} (a_1 b_1) \tilde{\mathbb{E}}'[C_i^{a_1} C_j^{a_2} C_k^{a_3} C_i^{b_1} C_p^{b_2}] \\
&= \sum_{a_1 b_1} (a_1 b_1) \tilde{\mathbb{E}}'[C_i^{a_1} C_i^{b_1}] \\
&= \tilde{\mathbb{E}}'[C_i C_i] \\
&= \tilde{\mathbb{E}}[x_i^2] \\
&= 1.
\end{aligned}$$

□

Acknowledgement. AWH and AN were funded by NSF Grant CCF-1629809 and AWH was funded by NSF grant CCF-1452616. XW was funded by the NSF Waterman Award of Scott Aaronson. All three authors (AWH, AN, and XW) were funded by ARO contract W911NF-12-1-0486. We are grateful to an anonymous STOC reviewer for pointing out a mistake in an earlier version.

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

References

[ABD+08] Aaronson, S., Beigi, S., Drucker, A., Fefferman, B., Shor, P.: The power of unentanglement. In: Annual IEEE Conference on Computational Complexity, pp. 223–236 (2008). arXiv:0804.0802

[ABD+09] Aaronson, S., Beigi, S., Drucker, A., Fefferman, B., Shor, P.: The power of unentanglement. Theory Comput. **5**(1), 1–42 (2009). arXiv:0804.0802

[AK08] Altunbulak, M., Klyachko, A.: The Pauli principle revisited. Commun. Math. Phys. **282**(2), 287–322 (2008). arXiv:0802.0918

[AS15] Aubrun, G., Szarek, S.J.: Dvoretzky's theorem and the complexity of entanglement detection (2015). arXiv:1510.00578

[Bar14] Barak, B.: Sum of squares upper bounds, lower bounds, and open questions (2014). https://www.boazbarak.org/sos/prev/files/all-notes.pdf. Accessed 26 Feb 2019

[BBH+12] Barak, B., Brandão, F.G.S.L., Harrow, A.W., Kelner, J., Steurer, D., Zhou, Y.: (2012) Hypercontractivity, sum-of-squares proofs, and their applications. In: Proceedings of the 44th Symposium on Theory of Computing, STOC '12, pp. 307–326. (2012). arXiv:1205.4484

[BBL+06] Brassard, G., Buhrman, H., Linden, N., Méthot, A.A., Tapp, A., Unger, F.: Limit on nonlocality in any world in which communication complexity is not trivial. Phys. Rev. Lett. **96**, 250401 (2006). arXiv:quant-ph/0508042

[BCHW15] Bäuml, S., Christandl, M., Horodecki, K., Winter, A.: Limitations on quantum key repeaters. Nat. Commun. **6**, 6908 (2015). arXiv:1402.5927

[BCY11] Brandão, F.G.S.L., Christandl, M., Yard, J.: Faithful squashed entanglement. Commun. Math. Phys. **306**(3), 805–830 (2011). arXiv:1010.1750

[Bei10] Beigi, S.: NP vs QMA_log(2). **10**(1&2), 0141–0151 (2010). arXiv:0810.5109

[Bel64] Bell, J.S.: On the Einstein–Podolsky–Rosen paradox. Physics **1**(3), 195–200 (1964)

[BFS15] Berta, M., Fawzi, O., Scholz, V.B.: Quantum bilinear optimization (2015). arXiv:1506.08810

[BGL17] Bhattiprolu, V., Guruswami, V., Lee, E.: Sum-of-squares certificates for maxima of random tensors on the sphere. In: LIPIcs-Leibniz International Proceedings in Informatics, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, vol. 81 (2017). arXiv:1605.00903

[BH13a] Brandão, F.G.S.L., Harrow, A.W.: Product-state approximations to quantum ground states. In: Proceedings of the 45th Annual ACM Symposium on Theory of Computing, STOC '13, pp. 871–880 (2013). arXiv:1310.0017

[BH13b] Brandão, F.G.S.L., Harrow, A.W.: Quantum de Finetti theorems under local measurements with applications. In: Proceedings of the 45th Annual ACM Symposium on Theory of Computing, STOC '13, pp. 861–870 (2013). arXiv:1210.6367

[BHK05] Barrett, J., Hardy, L., Kent, A.: No signaling and quantum key distribution. Phys. Rev. Lett. **95**, 010503 (2005). arXiv:quant-ph/0405101

[BKS14] Barak, B., Kelner, J.A., Steurer, D.: Rounding sum-of-squares relaxations. In: Proceedings of the 46th Annual ACM Symposium on Theory of Computing, STOC '14, pp. 31–40. ACM (2014). arXiv:1312.6652

[BP15] Bamps, C., Pironio, S.: Sum-of-squares decompositions for a family of Clauser–Horne–Shimony–Holt-like inequalities and their application to self-testing. Phys. Rev. A. **91**, 052111 (2015). arXiv:1504.06960

[BPR16] Braun, G., Pokutta, S., Roy, A.: Strong reductions for extended formulations (2016). arXiv:1512.04932

[BPZ15] Braun, G., Pokutta, S., Zink, D.: Inapproximability of combinatorial problems via small LPs and SDPs. In: Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC '15, pp. 107–116. ACM, New York (2015)

[BS10] Beigi, S., Shor, P.W.: Approximating the set of separable states using the positive partial transpose test. J. Math. Phys. **51**(4), 042202 (2010). arXiv:0902.1806

[BS16] Barak, B., Steurer, D.: Sum-of-squares proofs and the quest toward optimal algorithms (2016). arXiv:1404.5236

[BT09] Blier, H., Tapp, A.: All languages in NP have very short quantum proofs. In: First International Conference on Quantum, Nano, and Micro Technologies, pp. 34–37. Los Alamitos, IEEE Computer Society (2009). arXiv:0709.0738

[BV04] Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)

[CB96] Chefles, A., Barnett, S.M.: Complementarity and Cirel'son's inequality. J. Phys. A: Math. Gen. **29**(10), L237 (1996)

[CD10] Chen, J., Drucker, A.: Short multi-prover quantum proofs for SAT without entangled measurements (2010). arXiv:1011.0716

[CFS02] Caves, C.M., Fuchs, C.A., Rüdiger, S.: Unknown quantum states: the quantum de Finetti representation. J. Math. Phys. **43**(9), 4537–4559 (2002). arXiv:quant-ph/0104088

[Cir80] Cirel'son, B.S.: Quantum generalizations of Bell's inequality. Lett. Math. Phys. **4**(2), 93–100 (1980)

[CJYZ15] Chen, J., Ji, Z., Yu, N., Zeng, B.: Detecting consistency of overlapping quantum marginals by separability (2015). arXiv:1509.06591

[CKMR07] Christandl, M., König, R., Mitchison, G., Renner, R.: One-and-a-half quantum de Finetti theorems. Commun. Math. Phys. **273**, 473–498 (2007). arXiv:quant-ph/0602130

[CM15] Chakraborty, R., Mazziotti, D.A.: Structure of the one-electron reduced density matrix from the generalized Pauli exclusion principle. Int. J. Quantum Chem. **115**(19), 1305–1310 (2015)

[CSW12] Christandl, M., Schuch, N., Winter, A.: Entanglement of the antisymmetric state. Commun. Math. Phys. **311**(2), 397–422 (2012). arXiv:0910.4151

[DF80] Diaconis, P., Freedman, D.: Finite exchangeable sequences. Ann. Probab. **8**, 745–764 (1980)

[DLTW08] Doherty, A.C., Liang, Y.-C., Toner, B., Wehner, S.: The quantum moment problem and bounds on entangled multi-prover games. In: CCC '08, pp. 199–210 (2008). arXiv:0803.4373

[DPD02] Terhal, B.M., DiVincenzo, D.P., Leung, D.W.: Quantum data hiding. IEEE Trans. Inf. Theory **48**(3), 580 (2002). arXiv:quant-ph/0103098

[DPS03] Doherty, A.C., Parrilo, P.A., Spedalieri, F.M.: A complete family of separability criteria. Phys. Rev. A. **69**, 022308 (2003). arXiv:quant-ph/0308032

[DPS04] Doherty, A.C., Parrilo, P.A., Spedalieri, F.M.: Complete family of separability criteria. Phys. Rev. A. **69**, 022308 (2004). arXiv:quant-ph/0308032

[DW12] Doherty, A.C., Wehner, S.: Convergence of SDP hierarchies for polynomial optimization on the hypersphere (2012). arXiv:1210.5048

[Gha10] Gharibian, S.: Strong NP-hardness of the quantum separability problem. QIC **10**(3&4), 343–360 (2010). arXiv:0810.4507

[Gol08] Goldreich, O.: Computational Complexity: A Conceptual Perspective, 1 edn.. Cambridge University Press, New York (2008)

[Gri01] Grigoriev, D.: Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. Theor. Comput. Sci. **259**, 613–622 (2001)

[Gur03] Gurvits L.: Classical deterministic complexity of Edmonds' problem and quantum entanglement (2003). arXiv:quant-ph/0303055

[HHH96] Horodecki, M., Horodecki, P., Horodecki, R.: Separability of mixed states: necessary and sufficient conditions. Phys. Lett. A. **223**(1–2), 1–8 (1996). arXiv:quant-ph/9605038

[HM04] Helton, W., McCullough, S.: A positivstellensatz for non-commutative polynomials. Trans. Am. Math. Soc. **356**(9), 3721–3737 (2004)

[HM13] Harrow, A.W., Montanaro, A.: Testing product states, quantum Merlin-Arthur games and tensor optimization. J. ACM **60**(1), 3–1343 (2013). arXiv:1001.0017

[HNW15] Harrow, A.W., Natarajan, A., Wu, X.: An improved semidefinite programming hierarchy for testing entanglement (2015). arXiv:1506.08834

[HP06] Helton, J.W., Putinar, M.: Positive polynomials in scalar and matrix variables, the spectral theorem and optimization (2006). arXiv:math/0612103

[IKM09] Ito, T., Kobayashi, H., Matsumoto, K.: Oracularization and two-prover one-round interactive proofs against nonlocal strategies. In: Proceedings: Twenty-Fourth Annual IEEE Conference on Computational Complexity (CCC 2009), pp. 217–228 (2009). arXiv:0810.0693

[IP01] Impagliazzo, R., Paturi, R.: On the complexity of k-SAT. J. Comput. Syst. Sci. **62**(2), 367–375 (2001)

[IV12] Ito, T., Vidick, T.: A multi-prover interactive proof for NEXP sound against entangled provers. In: FOCS '12 (2012). arXiv:1207.0550

[Kho02] Khot, S.: On the power of unique 2-prover 1-round games. In: STOC, pp. 767–775 (2002)

[KKM+11] Kempe, J., Kobayashi, H., Matsumoto, K., Toner, B., Vidick, T.: Entangled games are hard to approximate. SIAM J. Comput. **40**(3), 848–877 (2011). arXiv:0704.2903

[Kly13] Klyachko, A.A.: The Pauli principle and magnetism (2013). arXiv:1311.5999

[KM09] Koenig, R., Mitchison, G.: A most compendious and facile quantum de Finetti theorem. J. Math. Phys. **50**, 012105 (2009). arXiv:quant-ph/0703210

[KMR17] Kothari, P.K., Meka, R., Raghavendra, P.: Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of CSPs. In: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, pp. 590–603. ACM (2017). arXiv:1610.02704

[KR05] Koenig, R., Renner, R.: A de Finetti representation for finite symmetric quantum states. J. Math. Phys. **46**(12), 122108 (2005). arXiv:quant-ph/0410229

[Lan15] Lancien C.: k-extendibility of high-dimensional bipartite quantum states (2015). arXiv:1504.06459

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

[Lau09] Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. Emerg. Appl. Algebraic Geom. **149**, 157–270 (2009)

[Liu07] Liu, Y.-K.: The complexity of the consistency and N-representability problems for quantum states. Ph.D. thesis, University of California, San Diego (2007). arXiv:0712.3041

[LNN12] Gall, F.L., Nakagawa, S., Nishimura, H.: On QMA protocols with two short quantum proofs. Quantum Inf. Comput. **12**, 0589 (2012). arXiv:1108.4306

[LRS15] Lee, J.R., Raghavendra, P., Steurer, D.: Lower bounds on the size of semidefinite programming relaxations. In: Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC '15, pp. 567–576. ACM, New York (2015). arXiv:1411.6317

[Mas06] Masanes, L.: All bipartite entangled states are useful for information processing. Phys. Rev. Lett. **96**, 150501 (2006). arXiv:quant-ph/0508071

[Maz04] Mazziotti, D.A.: Realization of quantum chemistry without wave functions through first-order semidefinite programming. Phys. Rev. Lett. **93**, 213001 (2004)

[MSW04] Matsumoto, K., Shimono, T., Winter, A.: Remarks on additivity of the Holevo channel capacity and of the entanglement of formation. Commun. Math. Phys. **246**, 427–442 (2004). arXiv:quant-ph/0206148

[MW12] Matthews, W., Wehner, S.: Finite blocklength converse bounds for quantum channels (2012). arXiv:1210.4722

[Nes00] Nesterov, Y.: Squared functional systems and optimization problems. High Perform. Optim. **13**, 405–440 (2000)

[NOP09] Navascues, M., Owari, M., Plenio, M.B.: The power of symmetric extensions for entanglement detection. Phys. Rev. A **80**, 052306 (2009). arXiv:0906.2731

[NPA08] Navascués, M., Pironio, S., Acin, A.: A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. New J. Phys. **10**(7), 073013 (2008). arXiv:0803.4290

[OWWZ14] O'Donnell, R., Wright, J., Wu, C., Zhou, Y.: Hardness of robust graph isomorphism, lasserre gaps, and asymmetry of random graphs. In: Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '14, pp. 1659–1677. SIAM (2014). arXiv:1401.2436

[Pal15] Palazuelos, C.: Random constructions in Bell inequalities: a survey (2015). arXiv:1502.02175

[Par00] Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Technical report, MIT, 2000. Ph.D Thesis

[Per96] Peres, A.: Separability criterion for density matrices. Phys. Rev. Lett. **77**(8), 1413–1415 (1996)

[Per12] Pereszlényi, A.: Multi-prover quantum Merlin–Arthur proof systems with small gap (2012). arXiv:1205.2761

[PH11] Poulin, D., Hastings, M.B.: Markov entropy decomposition: a variational dual for quantum belief propagation. Phys. Rev. Lett. **106**(8), 80403 (2011). arXiv:1012.2050

[PNA10] Pironio, S., Navascués, M., Acín, A.: Convergent relaxations of polynomial optimization problems with noncommuting variables. SIAM J. Optim. **20**(5), 2157–2180 (2010). arXiv:0903.4368

[PY91] Papadimitriou, C.H., Yannakakis, M.: Optimization, approximation, and complexity classes. J. Comput. Syst. Sci. **43**, 425–440 (1991)

[Rai01] Rains, E.M.: A semidefinite program for distillable entanglement. IEEE Trans. Inf. Theory, **47**(7), 2921–2933 (2001). arXiv:quant-ph/0008047

[RST12] Raghavendra, P., Steurer, D., Tulsiani, M.: Reductions between expansion problems. In: CCC, pp. 64–73 (2012). arXiv:1011.2586

[Rus07] Mary Beth, R.: Connecting n-representability to Weyl's problem: the one-particle density matrix for $\mathrm{N} = 3$ and $\mathrm{R} = 6$. J. Phys. A: Math. Theor. **40**(45), F961 (2007). arXiv:0706.1855

[RUV13a] Reichardt, B.W., Unger, F., Vazirani, U.: Classical command of quantum systems. Nature **496**(7446), 456–460 (2013). arXiv:1209.0449

[RUV13b] Reichardt, B.W., Unger, F., Vazirani, U.: A classical leash for a quantum system: command of quantum systems via rigidity of chsh games. In: Proceedings of the 4th Conference on Innovations in Theoretical Computer Science, ITCS '13, pp. 321–322 (2013). arXiv:1209.0448

[Sch08] Schoenebeck, G.: Linear level Lasserre lower bounds for certain k-CSPs. In: Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS '08, pp. 593–602. IEEE Computer Society, Washington, DC (2008)

[Sho87] Shor, N.Z.: An approach to obtaining global extremums in polynomial mathematical programming problems. Cybern. Syst. Anal. **23**(5), 695–700 (1987)

[Slo16] Slofstra, W.: Tsirelson's problem and an embedding theorem for groups arising from non-local games (2016). arXiv:1606.03140

[SW08] Scholz, V.B., Werner, R.F.: Tsirelson's problem (2008). arXiv:0812.4305

[Tre12] Trevisan, L.: On Khot's unique games conjecture. Bull. AMS **49**(1), 91–111 (2012)

[Tul09] Tulsiani, Madhur: CSP gaps and reductions in the Lasserre hierarchy. In: Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing, pp. 303–312. ACM (2009)

[Vid13] Vidick, T.: Three-player entangled XOR games are NP-hard to approximate. In: Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, FOCS '13, pp. 766–775. IEEE Computer Society (2013). arXiv:1302.1242

Communicated by M. M. Wolf