Limitations of monogamy, Tsirelson-type bounds, and other semidefinite programs in quantum information

Aram W. Harrow∗ Anand Natarajan∗ Xiaodi Wu†

January 26, 2016

Abstract

We introduce a method for using reductions to prove limitations on the ability of semidefinite programs (SDPs) to approximately solve optimization problems. We use this to show specifically that SDPs have limited ability to approximate two particularly important sets in quantum information theory:

1. The set of separable (i.e. unentangled) states.
2. The set of quantum correlations; i.e. conditional probability distributions achievable with local measurements on a shared entangled state.

The former result implies a proof of a version of the ”no approximate disentangler” conjecture of Watrous and a no-go theorem for SDP relaxations of the 2 → 4 norm of a matrix, while the latter is the first unconditional negative result for SDP relaxations of noncommutative polynomial optimization. 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) and show that any SDPs which approximately solve either of these problems must have a number of variables which grows very quickly with problem size.

These results can be viewed as limitations on various ideas from quantum information (including the monogamy principle, the PPT test, Tsirelson-type inequalities) for understanding entanglement. Many of these ideas have been formalized into SDP hierarchies by Doherty-Parrilo-Spedalieri, Navascues-Pironio-Acin and Berta-Fawzi-Scholz, all of which we establish limits on the effectiveness of.

Our techniques are based on reductions between integrality gaps. This allows the recent integrality gaps of Lee-Raghavendra-Steurer to be extended to a wide range of problems, without needing to redo their use of Fourier analysis to non-boolean domains.

∗Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
†Department of Computer and Information Science, University of Oregon, Eugene, OR 97403, USA.
1 Introduction

1.1 Background and Motivation

1.1.1 Semidefinite Programming (SDP) and Sum-of-Squares (SoS) hierarchies

A major question in the theory of algorithms and complexity is the power of Semidefinite Programming (SDP) relaxations [BV04]. Most problems in NP admit various SDP relaxations, but worst-case quality of the resulting approximations is often unknown. One particularly promising family of SDP relaxations is the sum-of-squares (SoS) hierarchy, introduced in [Sho87, Nes00, Par00, Las01] and reviewed in [Lau09, Bar14]. The SoS hierarchy 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. If $k$ needs to be $O(n)$ then this will in general be no better than brute-force search, but in some cases 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].

1.1.2 Polynomial optimization and unentangled quantum states

One of the main motivations of this paper is to understand the difficulty of polynomial optimization, meaning the problem of maximizing an $O(1)$-degree polynomial in $n$ variables subject to polynomial equality or inequality constraints. This problem is roughly equivalent to a variety of other hard optimization problems [BBH+12, HM13], including determining whether a mixed quantum state is entangled or not, estimating the $2 \to 4$ norm of a matrix or estimating the minimum output Rényi entropy of a quantum channel. For concreteness we will focus on the optimization problem $h_{\text{Sep}(d,d)}$, defined for a positive-semidefinite $d^2 \times d^2$ matrix $M$ as

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

In general these problems cannot be approximated in polynomial time even to constant error [BBH+12, HM13], if we assume the Exponential Time Hypothesis (ETH) [IP01], which asserts that 3-SAT instances of size $n$ require $2^{\Omega(n)}$ time to solve.

Despite these worst-case hardness results (which we will discuss further in Section 1.2), a large body of work has been devoted to answering these questions in specific, often operationally useful, contexts. 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].

Given the utility of solving these problems even in specific cases (despite their worst-case hardness), a variety of heuristics have been developed, which we briefly outline here. 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 some entangled
states are PPT while some are not. Doherty, Parrilo and Spedalieri improved this to a hierarchy of approximations \[\text{DPS04}\]. The \(k\)th level of the so-called DPS hierarchy approximates the set of entangled states by the set of states \(\rho^{AB}\) for which there does not exist \(\tilde{\rho}^{A_1...A_k B_1...B_k}\) with \(\rho^{AB} = \tilde{\rho}^{A_1 B_1}\), the supports of \(\tilde{\rho}^{A_1...A_k}\) and \(\tilde{\rho}^{B_1...B_k}\) 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. 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. 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 counterexample showing that this weaker hierarchy still makes large errors until \(k > d/2\). This can also be turned into a counterexample 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 counterexample 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 \(k = O(\log d)\) \[\text{BCY11}\]

All of these counterexamples are 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 \[\text{BS10}\] 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 \(k\) \[\text{HNW15}\].

1.1.3 Noncommutative Polynomial Optimization and Quantum games

A second major class of optimization problems concerns polynomials in non-commuting variables \[\text{HP06 PNA10}\]. As we explain in Section 2.2, 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 state \[\text{NPA08}\]. The most famous example of a quantum-but-not-classical correlation was discovered by Bell in 1964 \[\text{Bel64}\] and gave a concrete experiment for which quantum 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 multiprover games where the provers share entanglement; 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 obtain smaller optimization problems, e.g. in quantum chemistry \[\text{Maz04}\].

A similar hierarchy was developed to approximate the set of correlations achievable with local measurements of quantum states \[\text{NPA08 DLTW08 BFS15}\], or for more general polynomial optimization \[\text{PNA10}\]. 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 \[\text{SW08}\]) whether it indeed converges to the correct value. Here too computational hardness results for this task are known (discussed in Section 1.2), but 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 communication complexity [BBL+06], to “self-testing” quantum systems and to multiparty secure quantum computing [BP13, RUV13] and to device-independent cryptography [BH05], to give a very partial list. 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].

1.2 Computational hardness

Most of the evidence against the power of the SoS and ncSoS hierarchies is based on hardness assumptions, such as \( P \neq NP \), the Unique Games Conjecture (UGC) or the Exponential Time Hypothesis (ETH).

To fix some notation, let \( \text{Sep}^k(d) \) denote the convex hull of \( |\psi_1\rangle \langle \psi_1| \otimes \cdots |\psi_k\rangle \langle \psi_k| \) as \( |\psi_1\rangle, \ldots, |\psi_k\rangle \) range over all unit vectors in \( \mathbb{C}^d \). For a general convex set \( S \), let \( h_S(x) := \max_{y \in S} \langle x, y \rangle \). Consider the problem of determining whether \( h_{\text{Sep}^k(d)}(M) \) 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 \( h_{\text{Sep}^k(d)} \) in this way.” We summarize these in Table 1.

<table>
<thead>
<tr>
<th>reference</th>
<th>( k )</th>
<th>( c )</th>
<th>( s )</th>
<th>( n )</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNN12</td>
<td>2</td>
<td>1</td>
<td>( 1 - \frac{1}{d \cdot \text{poly log}(d)} )</td>
<td>( O(d) )</td>
<td>(1)</td>
</tr>
<tr>
<td>Per12</td>
<td>2</td>
<td>1</td>
<td>( 1 - \frac{1}{\text{poly}(d)} )</td>
<td>( O(d) )</td>
<td>(2)</td>
</tr>
<tr>
<td>ABD+09</td>
<td>( \sqrt{d} \cdot \text{poly log}(d) )</td>
<td>( 0.99 )</td>
<td>( O(d) )</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td>CD10</td>
<td>( \sqrt{d} \cdot \text{poly log}(d) )</td>
<td>( 1 - 2^{-d} )</td>
<td>( 0.99 )</td>
<td>( O(d) )</td>
<td>(4)</td>
</tr>
<tr>
<td>HM13</td>
<td>2</td>
<td>1</td>
<td>( 0.01 )</td>
<td>( \frac{\log^2(d)}{\text{poly log}(d)} )</td>
<td>(5)</td>
</tr>
</tbody>
</table>

Table 1: Hardness results for \( h_{\text{Sep}^k(d)} \) (see Section 1.2). Notes: (1) This builds on work in [Gur03, BT09, Bei10] which achieved the same result with \( s = 1 - 1/\text{poly}(d) \). Related results were found for testing membership in \( \text{Sep}^2(d) \) in [Gur03, Liu07, Gha10]. (2) The measurement \( M \) can be implemented using a uniform quantum circuit of size \( \text{poly log}(d) \). (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 \( A_i, B_i \geq 0 \).

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 \( h_{\text{Sep}} \), and in particular implies the existence of integrality gaps for the SoS hierarchy. One hardness result that does not fit into this framework is the result by [BBH+12] that a constant-factor multiplicative approximation to \( h_{\text{Sep}^2(d)} \) could be used to solve Unique Games instances of size \( d^{O(1)} \).
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.

<table>
<thead>
<tr>
<th>reference</th>
<th>$k$</th>
<th>$c$</th>
<th>$s$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[KKM+11]</td>
<td>3</td>
<td>1</td>
<td>$1 - \frac{1}{\text{poly}(Q)}$</td>
<td>$O(Q)$</td>
</tr>
<tr>
<td>[IKM09]</td>
<td>2</td>
<td>1</td>
<td>$1 - \frac{1}{\text{poly}(Q)}$</td>
<td>$O(Q)$</td>
</tr>
<tr>
<td>[IV12]</td>
<td>4</td>
<td>1</td>
<td>$2^{-Q^{\Omega(1)}}$</td>
<td>$Q^{\Omega(1)}$</td>
</tr>
<tr>
<td>[Vid13]</td>
<td>3</td>
<td>1</td>
<td>$2^{-Q^{\Omega(1)}}$</td>
<td>$Q^{\Omega(1)}$</td>
</tr>
</tbody>
</table>

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

1.3 Main Results

In this paper we describe instances on which the above hierarchies (and others) fail to give the correct answer and thus provide unconditional lower bounds. Our results can be viewed as an extension of the recent celebrated progress of Lee, Raghavendra, and Steurer [LRS14] on the boolean cube to much more general domains.

On the set of separable states. Our results show that with no computational assumption the DPS hierarchy cannot estimate (1.1) to constant accuracy at level $k$ unless $k \geq \tilde{\Omega}(\log d)$, corresponding to run-time $d^{\tilde{\Omega}(\log d)}$ (see Corollary 4.10). In fact, we prove the following much more general statement,

**Theorem 1.1 (Informal, refer to Theorem 4.9)**. Any SDP relaxation of $h_{\text{Sep}}$ achieving constant accuracy the total number of variables must be $\geq d^{\tilde{\Omega}(\log d)}$.

Here “relaxation” is a technical condition defined in the body of our paper; roughly speaking a relaxation should replace the optimization over separable states with an optimization over a convex superset. (In fact we rule out a slightly larger class of approximations, and the ”relaxation” condition is from [LRS14].) This yields the first unconditional quantitative limits to the rate at which the DPS hierarchy converges.

A number of other computational results are known for $h_{\text{Sep}}$, covering cases such as $1/\text{poly}(d)$ accuracy or multipartite variants, and in some cases with known algorithms nearly matching computational hardness. In each of these we achieve unconditional no-go theorems for SDPs roughly matching the previously known ETH-based hardness. (See Section 4.4) These are not only the first limitations on DPS convergence known but (as with the ETH-based hardness results) they are already nearly tight in some cases: specifically the case of $\tilde{O}(1/d)$ accuracy [LNN12, NOP09] and constant-error many-party states with Bell measurements [CD10, BH13].

Our results have also led to seemingly unrelated corollaries:

- One corollary of these results (Theorem 4.13) is an unconditional proof of a version of the “no approximate disentangler” conjecture of Watrous, 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 $\text{Sep}(d, d) \approx \text{Image}(\mathcal{N})$ (with “$\approx$” defined precisely in the body of our paper, but roughly speaking it corresponds to constant error) we must have $D \geq d^{\omega(1)}$. 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 \cite{BCY11} but is still an open question for trace distance. The original goal of the conjecture was to rule out a particular strategy for putting QMA(2) inside QMA.\footnote{Here QMA(2) is the set of languages where membership can be verified using a proof that is a pair of unentangled quantum states.} Our results imply unconditional lower bounds on the input dimension $D$ with again nearly the same parameters previously known assuming ETH.

- Another corollary (Corollary 4.11) rules out small SDP relaxations for the $2 \rightarrow 4$ norm of a matrix, a problem with close connections to the Unique Games Conjecture \cite{BBH+12}. Here too no integrality gaps were known for $\omega(1)$ levels of the SoS hierarchy. For the full statement, see Corollary 5.11.

**On the set of quantum correlations.** As before we also give dimension lower bounds for any SDP relaxation achieving reasonable accuracy.

**Theorem 1.2 (Informal, refer to Theorem 5.8)** There exists a sequence of two-player non-local games $G_n$ such that any SDP relaxation approximating the entangled game value $w_{\text{entangled}}(G_n)$ to precision $O(1/n^2)$ has size $r_n \geq (n/\log n)^\Omega(n)$.

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

**Theorem 1.3 (Informal, refer to Theorem 5.10)** There exists a sequence of two-player non-local games $G_n$ such that the entangled game value $w_{\text{entangled}}(G_n) \leq 1 - c/n^2$ for some constant $c$ but the ncSOS hierarchy believes $w_{\text{entangled}}(G_n) = 1$ up to level $m = \Omega(n)$.

Previously we could not (unconditionally) exclude the possibility that the ncSoS hierarchy gave the exactly correct answer for some constant level $k$, and the previous examples gave no hint of how $k$ should scale with the accuracy or the number of questions.

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. As with the case of $h_{\text{Sep}}$, our results extend to general SDPs for quantum correlations and in particular also limit the stronger hierarchy described in \cite{BFS15}.

### 1.4 Technical Contributions

Ultimately, all of our hardness results stem from a classic result of Grigoriev \cite{Gri01}, showing hardness for the problem 3XOR in the SoS model, together with a recent breakthrough by Lee, Raghavendra and Steurer \cite{LRS14}, generalizing SoS results to general SDP lower bounds. To obtain integrality gaps for quantum problems considered here, we need to reduce (and sometimes embed) the classical hard problems in the SoS model to quantum ones. There have been quite a few examples using reductions to prove the hardness in the SoS model (e.g., \cite{Tul09, OWWZ14}). However, each proof requires slightly different properties about reductions and there is no explicit unified framework for doing so. We formulate a framework of reductions for the purpose of establishing
integrality gaps in Section 3, which aims to serve as one such framework to facilitate the proof of hardness in both the SoS model and the more general SDP model.

To obtain integrality gaps in the SoS model, one needs to show that (a) the SoS solution believes the value is large up to very high level (degree), and (b) the true value is actually very small. To achieve (a), we invent 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. To achieve (b), we need the reductions to have some kind of soundness. To both ends, we observe that known protocols in quantum interactive proofs could serve as the reductions. Specifically, we use the QMA(2) protocol for solving 3-SAT problem [ABD+08] for the hardness on the set of separable states and the two-player non-local game protocol in deriving the NP-hardness for these games [IKM09] for the hardness on the set of quantum correlations. We explicitly prove that these reductions are low-degree and the soundness is roughly due to the soundness of these quantum interactive protocols.

It is much more challenging to obtain integrality gaps for our problems in the SDP model because [LRS14] only works on the boolean cube \(\{0,1\}^n\), while the domain of separable states is hypersphere and the domain of approximating quantum non-local correlations is even the set of all quantum strategies over infinite-dimensional Hilbert spaces. In particular, the proof of [LRS14] made intricate use of the structure of \(\{0,1\}^n\), e.g. how functions on it are affected by noise, and it appeared initially that to extend their results to other domains (such as the sphere) would require a similar analysis in each new setting (e.g. considering spherical harmonics). Moreover, the SDP lower bound in [LRS14] has a technical restriction on the type of SDP relaxations, which we formulate as the embedding property. This technical restriction puts additional difficulty in constructing the reductions.

Our crucial observation is that these quantum interactive proof protocols actually allow one to embed hard problems over \(\{0,1\}^n\) into larger, even infinite-dimensional, domains. Thus, we can avoid extending [LRS14]’s analysis to each setting, while still extend its results on \(\{0,1\}^n\) to much larger domains. Our strategy is quite general and we hope it can find other extensions of [LRS14].

As our last technical contribution, we also demonstrate how to derive a ncSoS solution from a SoS solution for the purpose of showing the hardness of approximating entangled non-local game values. Our idea is to embed a SoS solution into a ncSoS solution, where the embedding is due to the connection between each question and prover’s corresponding strategy (i.e., operators). An interesting observation is that 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 summarize the complete reductions that handle all technical details in Figure 1.

1.5 Open Problems

Our work leaves open a number of intriguing questions.

- It raises the motivation to examine convex but non-SDP-based relaxations, such as the entropic bounds used in [PH11].

---

^2This roughly refers to the ”Vector Completeness” in [Lrs14] and ”SoS Completeness” in [Owwz14]. It explicitly requires the existence of a mapping that is a polynomial of low-degree, which maps a SoS solution of the original problem to a SoS solution of the reduced one.
• 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] do not rely on the probabilistic method.

• 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.

• Our results can be viewed as an extension of [LRS14] beyond $\{0,1\}^n$ to sets such as the hypersphere. We believe that finding further extensions of [LRS14] is a promising avenue.

2 Preliminaries

We summarize relevant background about quantum information and our terminology in Section 2.1. We provide a brief introduction to the sum-of-squares proof/optimization in Section 2.2.
2.1 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 $\rho$, i.e., a positive semidefinite matrix with trace 1, over $\mathcal{A}$. The set of all quantum states in $\mathcal{A}$ is denoted by $\text{Dens}(\mathcal{A})$. A quantum state $\rho$ is called a pure state if $\text{rank}(\rho) = 1$; otherwise, $\rho$ is called a mixed state. An $m$-qubit pure state $\rho = |\psi\rangle\langle\psi|$ 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 $L(\mathcal{A})$ is defined by $\langle X, Y \rangle = \text{tr}(X^\dagger Y)$ for all $X, Y \in L(\mathcal{A})$, where $\dagger$ is the adjoint operator.

Let $\Sigma$ be a finite nonempty set of measurement outcomes. A positive-operator valued measure (POVM) on the state space $\mathcal{A}$ with outcomes in $\Sigma$ is a collection of positive semidefinite operators $\{P_a : a \in \Sigma\}$ such that $\sum_{a \in \Sigma} P_a = \text{id}_\mathcal{A}$. 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 $\rho$, the probability of each outcome $a \in \Sigma$ is $\langle \rho, P_a \rangle$. When outcome $a$ is observed, the quantum state $\rho$ becomes the state $\sqrt{P_a} \rho \sqrt{P_a} / \langle \rho, P_a \rangle$.

Distance Measures. For any $X \in L(\mathcal{A})$ with singular values $\sigma_1, \ldots, \sigma_d$, where $d = \text{dim}(\mathcal{A})$, the trace norm of $\mathcal{A}$ is $\|X\|_{\text{tr}} = \sum_{i=1}^d \sigma_i$. The trace distance between two quantum states $\rho_0$ and $\rho_1$ is defined to be

$$\|\rho_0 - \rho_1\|_{\text{tr}} \overset{\text{def}}{=} \frac{1}{2} \|\rho_0 - \rho_1\|_{\text{tr}}.$$

2.2 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].

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. The non-commutative 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 $\bigoplus_{d' \leq d} \text{Sym}^{d'} \mathbb{R}^n$, where $\text{Sym}^{d'} V$ denotes the symmetric subspace of $V^{\otimes d'}$.

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

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

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

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

Note that the supremum here is over density operators $\rho$ 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 $f_{\text{max}}$ or $F_{\text{max}}$ from above. This complements simply guessing values of $x$ or $(\rho, X)$ which provides lower bounds on $f_{\text{max}}$ or $F_{\text{max}}$ when they satisfy the constraints. 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

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

Observe that the RHS is $\geq 0$ 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, 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 $n^{O(d) + O(1)}$ using semidefinite programming \cite{Lau09}.

If we find the minimum $c$ for which (2.3) holds, then we obtain a hierarchy of upper bounds on $f_{\text{max}}^d$, referred to as the SoS hierarchy or the Lasserre hierarchy. Denote this upper bound by $F_{\text{SoS}}^d$. Given mild assumptions on the constraints $g_1, \ldots, g_m$ one can prove that $\lim_{d \to \infty} F_{\text{SoS}}^d = f_{\text{max}}$ \cite{Lau09}. The tradeoff between degree $d$ and error ($F_{\text{SoS}}^d - f_{\text{max}}$) 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

$$c - F = \sum_{i=1}^{k} P_i^* P_i + \sum_{i=1}^{m} Q_i G_i R_i,$$  \hspace{1cm} (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 $n^{O(d) + O(1)}$, and we denote the corresponding value by $F_{\text{ncSoS}}^d$. It is known that $F_{\text{max}} \leq F_{\text{ncSoS}}^d$ for all $d$ and $\lim_{d \to \infty} F_{\text{ncSoS}}^d = F_{\text{max}}$ \cite{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 $\tilde{E}$ is an element of $\mathbb{R}[x]_d^*$ (i.e. a linear map from $\mathbb{R}[x]_d$ to $\mathbb{R}$) satisfying

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

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

$$f_{\text{SoS}}^d = \max \{ \tilde{E}[f] : \tilde{E} \text{ is a degree-$d$ pseudo-expectation satisfying } g_1, \ldots, g_m \}.  \hspace{1cm} (2.5)$$

The term “pseudo-expectation” comes from the fact that for any distribution $\mu$ over $R^n$ we can define a pseudo-expectation $\tilde{E}[f] := \mathbb{E}_{x \sim \mu}[f(x)]$. Thus the set of pseudo-expectations can be thought of as the low-order moments that could come from a “true” distribution $\mu$ or could come from a “fake” distribution. Indeed an alternate approach (which we will not use) 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 \( \leq d \) [LRS14]. We can define a noncommutative pseudo-expectation \( \tilde{E} \in \mathbb{R} \langle X \rangle^d \) similarly by the constraints \( \tilde{E}[1] = 1 \) and \( \tilde{E}[p^2/p] \geq 0 \) 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 \( \tilde{E} \) is a degree-\( d \) pseudo-expectation over \( \{\pm 1\}^n \) if for any variable \( x_i \) and polynomial \( q \) of degree at most \( d - 2 \),

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

(2.6)

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

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 relaxations 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 \( \Delta^A = \{\Delta^A_n\}_{n \in \mathbb{N}} \), is a family of collections of optimization instances that are parameterized by \( n \in \mathbb{N} \) and which consists of following components:

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

\[
\text{OPT}(\Phi^A_n) := \max_{x \in \mathcal{P}^A_n} \Phi^A_n(x).
\]

An example of an optimization problem is MAX-CUT, in which \( n \) is the number of vertices in a graph, \( \mathcal{P}^A_n = \{0,1\}^n \) and \( \Delta^A_n \) is the set of functions of the form \( \Phi^A_n(x) := E_{(i,j) \sim E}(x_i - x_j)^2 \) for some \( E \subset [n] \times [n] \). As we can see from the example, the functions \( \Phi^A_n \) 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 \( P_n \) is a variety of \( \mathbb{R}^m \) for some bounded function \( m = m(n) \) and any instance \( \Phi^n_A : P^n_A \to [0, 1] \) is a polynomial. Here “variety” means that

\[
P_n = \{ x \in \mathbb{R}^m : g_1(x) = \cdots = g_m(x) = 0 \},
\]

for some polynomials \( g_1, \ldots, g_m \).

Definition 3.3 (Boolean Polynomial Optimization) A boolean polynomial optimization problem \( A \), denoted by \( \Pi^n_A \), is a polynomial optimization problem defined by the constraints \( x_i^2 = 1 \) for \( i = 1, \ldots, n \). Thus the feasible set \( P^n_A \) is the boolean hypercube \( \{ \pm 1 \}^n \).

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 constraints \( \{ g_1, \ldots, g_m \} \), where each constraint \( g_i : \{ \pm 1 \}^n \to \{ 0, 1 \} \) is a boolean function, and the objective function \( f = \frac{1}{m} \sum_{i=1}^m g_i \) counts the fraction of clauses that evaluate to 1.

Proposition 3.5 Any CSP where each constraint depends on \( \kappa \) variables can be written as a boolean polynomial optimization problem, where the objective function is a polynomial of degree \( 2\kappa \).

Proof. We first show that each constraint can be expressed as a low-degree polynomial. Given a string \( (y_1, \ldots, y_k) \in \{ \pm 1 \}^\kappa \), we define the indicator function

\[
1_{y_1, \ldots, y_k}(x_1, \ldots, x_\kappa) = \prod_{i=1}^\kappa (1 - \frac{1}{4}(x_i - y_i)^2).
\]

This function is a polynomial of degree \( 2\kappa \). It is easy to see that when \( x_i = y_i \) for all \( i \), then \( 1_{y_1, \ldots, y_k}(x_1, \ldots, x_\kappa) = 1 \), and if \( x_i \neq y_i \) for any \( i \), then \( 1_{y_1, \ldots, y_k}(x_1, \ldots, x_\kappa) = 0 \). Using these indicator functions, we can express any boolean function \( g \) over \( \kappa \) variables as a polynomial with degree \( 2\kappa \):

\[
g(x_1, \ldots, x_\kappa) = \sum_{(y_1, \ldots, y_\kappa) \in \{ \pm 1 \}^\kappa} g(y_1, \ldots, y_\kappa) 1_{y_1, \ldots, y_\kappa}(x_1, \ldots, x_\kappa).
\]

Thus, if we are given a CSP with constraints \( \{ g_1, \ldots, g_m \} \), each of which depends on \( \kappa \) variables, then the total objective function \( f = \frac{1}{m} \sum_{i=1}^m g_i \) is a polynomial of degree \( 2\kappa \).

Another important class of optimization problems are operator norms of linear functions, defined as follows. If \( A, B \) are normed spaces and \( T : A \to B \) is a linear map then \( \| T \|_{A \to B} := \sup_{x \neq 0} \| T(x) \|_B / \| x \|_A \) can be thought of as an optimization problem where \( P_n \) is the unit ball of \( A \) and \( \Phi^n_A(x) = \| T(x) \|_B \). If \( A = \ell_p^n \) and \( B = \ell_q^m \) then this corresponds to a polynomial optimization problem. Computing \( h_{\text{sep}} \) (cf. \([1, 1]\)) can be similarly be formulated as a polynomial optimization problem, where the feasible set is the unit sphere \([\text{BBH} + 12]\).
Our goal is to find optimization problems where the SoS hierarchy and other SDP relaxations 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 relaxation. 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 $\Phi_m^A \in \Delta_m^A$ be an instance of optimization $A$ for some $m$. A degree-$d$ value-$c$ pseudo-solution for $\Phi_m^A$ is a degree-$d$ pseudo-expectation $\tilde{E}$ satisfying the constraints of $P_n^A$ such that

$$\tilde{E}[\Phi_m^A(x)] \geq c$$

In the case of CSPs, we can also define a stronger type of pseudo-solution that not only achieves a high objective value, but also satisfies the constraints of the CSP:

**Definition 3.7** Given a CSP $A$ whose objective function is a polynomial of degree $k$, we say a degree-$d$ pseudo-expectation $\tilde{E}[\cdot]$ perfectly satisfies $A$ if for every constrain $g_i(x)$ of $A$, and every polynomial $p(x)$ with $\deg(p) \leq d - 2k$,

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

A single degree-$d$ value-$c$ pseudo-solution for an instance $\Phi_m^A$ implies the sum-of-squares approach (up to degree $d$) believes the optimum value of $\Phi_m^A$ is at least $c$. If the true optimum value of $\Phi_m^A$ 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 $m$.

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

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

**Proposition 3.9** Let $A$ be any polynomial optimization problem with a degree-$d$ value-$(c, s)$ integrality gap. Let $\delta = \frac{1}{2}(c - s)$ and $f_n = c - \delta - \Phi_n^A(x_n^A)$ where $\Phi_n^A$ is from the integrality gap and $x_n^A \in P_n^A$. Thus $f_n$ is a polynomial over $P_n^A$ and has $\deg_{\text{SOS}}(f_n) \geq d$.

**Proof.** The proof follows directly by definition.  

12
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 by 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 $\Delta^A$ to $\Delta^B$; i.e. $R(\Phi^A_n) \in \Delta^B_n$.

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** if for any $n$ and any $\Phi^A_n$ and its corresponding $\Phi^B_n = R(\Phi^A_n)$, we have
  $$\text{OPT}(\Phi^A_n) = \max_{x \in P^A_n} \Phi^A_n(x) \leq s^A \Rightarrow \text{OPT}(\Phi^B_n) = \max_{x \in P^B_n} \Phi^B_n(x) \leq s^B.$$  
  Here $s^A, s^B$ are understood to be functions of $n$.

- **embedded** if for any $n$, there is an additional map $E : P^A_n \mapsto P^B_n$ such that for any $\Phi^A_n$ and its corresponding $\Phi^B_n$, any $x^A_n \in P^A_n$ and its corresponding $x^B_n = E(x^A_n) \in P^B_n$, we have
  $$\Phi^A_n(x^A_n) = \Phi^B_n(x^B_n).$$

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 $x \in P^A_n$ (i.e. including non-optimal $x$) corresponds to some point in $P^B_n$ with the same value under $\Phi^B_n$. This condition was needed for the recent SDP lower bounds in [LRS14].

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

**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 $\Phi^A_n$, there is a degree-$d^B$ value-$c^B$ pseudo-solution for its corresponding instance $\Phi^B_n$, for any $n$. Here $d^A, c^A, d^B, c^B$ should be thought of as functions of $n$.

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 $R_{B \Rightarrow C}$. 

13
• 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. It follows directly by definition.

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-$d$ 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 construct a pseudo-solution from another under a polynomial map in the following sense. We say that $p : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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

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

(3.1a)

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

(3.1b)

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

Suppose that $p$ is a degree-$d$ polynomial map from $\mathbb{R}^n \rightarrow \mathbb{R}^m$ such that $p(A) \subseteq B$. Let $E_A \in \mathbb{R}[x_1, \ldots, x_n]_{\ell/d}$ be a degree-$\ell$ 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-$\ell/d$ pseudo-expectation $E_B \in \mathbb{R}[y_1, \ldots, y_m]_{\ell/d}$ 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-$\ell$ pseudo-density on $A$ is a function $\mu_A : A \rightarrow \mathbb{R}$ 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 $f \in \mathbb{R}[x]_{\ell/2}$. 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-$\ell = 2n$ pseudo-densities are also probability distributions. In general a degree-$\ell$ pseudo-density $\mu_A$ induces a pseudo-expectation $E_A \in \mathbb{R}[x]_{\ell}$ with

$$E_A[f] := \sum_{x \in A} \mu_A(x)f(x),$$

(3.2)

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

14
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 \( \mu_A \) as a linear map from \( \mathbb{R}[x] \to \mathbb{R} \). Given a degree-\( \ell \) pseudo-expectation \( \tilde{E}_A \), (3.2) can be thought of as constraining \( \mu_A \) 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 \( \mu_A \) 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.

By the above arguments, \( \tilde{E}_B \in \mathbb{R}[y]_{\ell/d} \) by
\[
\tilde{E}_B[f] := \sum_{y \in B} \mu_B(y)f(y).
\]

By the above arguments, \( \tilde{E}_B[1] = 1 \) and \( \tilde{E}_B[f^2] \geq 0 \) 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
\[
\tilde{E}_B[h_iq] = \sum_{y \in B} \mu_B(y)h_i(y)q(y) = 0,
\]
since \( h_i(B) = 0 \). Thus \( \tilde{E}_B \) 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 \( f : P_n^A \to P_n^B \), such that if \( x_A \in P_n^A \) satisfies all the constraints of an instance of \( A \), then \( f(x_A) \in P_n^B \) satisfies all the constraints of the corresponding instance of \( B \),
- each coordinate of \( f(x_A) \) depends on at most \( \kappa \) coordinates of \( x_A \), and
- there exists a degree-\( d^A \) pseudo-solution that perfectly satisfies \( A \).

Then there exists a degree-\( d^A/2\kappa \) 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 \( 2\kappa \). Now, suppose that the source problem \( A \) has a perfectly satisfying degree-\( d^A \) pseudo-solution, given by a pseudo-expectation operator \( \tilde{E}_A^A[\cdot] \). Then lemma applied to the degree-\( 2\kappa \) polynomial maps constructed above, yields a degree-\( d^A/2\kappa \) perfectly satisfying pseudo-solution for \( B \).
3.3 General SDP lower bounds

In this section we illustrate the recent celebrated result of Lee, Raghavendra, and Steurer (LRS) [LRS14] on the lower bounds of any semidefinite programming (SDP) relaxations 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 LRS result plays a crucial role in proving the our main SDP lower bounds in this paper. Our contribution can be viewed as developing techniques to apply LRS to general optimization problems.

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

**Definition 3.17 (SDP)** A semidefinite program (SDP) $A$ is an optimization problem with following restrictions: for each $n \in \mathbb{N}$,

- The feasible set $P_A^n$ 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 $\Phi^n_A$ is an affine function from $L(\mathbb{R}^r)$ to $[0,1]$.

**Definition 3.18 (SDP relaxation)** For any optimization problem $A$, a SDP $B$ is called $(s^B, s^A)$-approximate relaxation of $A$ if there exists an embedded reduction $R_{A \Rightarrow B}$ that is $(s^B, s^A)$-approximate.

We note that the SDP relaxation defined above is a more stringent concept than conventional ones, both because of the embedding property and because we require that the constraints do not depend on the objective function.

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

**Proposition 3.19** Let a SDP $C$ be a $(s^C, s^B)$-approximate SDP relaxation 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 relaxation 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. □

For any boolean polynomial optimization problem $\Pi^A$, the LRS result lower bounds the size of any reasonable SDP relaxation given some integrality gap for $\Pi^A$. We state the main result of LRS in a slightly more general way and leave the explanation how LRS result implies this statement in Appendix A.1.

**Proposition 3.20 (LRS)** Let $\Pi^A$ be any boolean polynomial optimization problem. If there is a degree-$d(n)$ value-$(c(n), s(n))$ integrality gap for $\Pi^A$, then any $(c([n/2]), s(n))$-approximate SDP relaxation of $\Pi^A$ must have size $r(n)$ lower bounded by

$$r(n) \geq \left( \frac{n}{\log n} \right)^{\Omega(d([n/2]))}.$$

As a direct consequence of the above proposition, we could extend the lower bound of SDP relaxation of $\Pi^A$ to any SDP relaxation of any problem $B$ in which there is an embedded reduction $R_{A \Rightarrow B}$ with matching approximation parameters by Proposition 3.19.
3.4 3XOR with integrality gap

In this section, we will introduce the source of all hardness we have for this paper, which is the 3XOR problem first discovered by Grigoriev [Gri01] and subsequently rediscovered by Schoenebeck [Sch08]. It is analogous to the prove 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 3.21 (3XOR)** 3XOR is a boolean polynomial optimization problem with the following restriction:

- **Instances**: for any $n$, an instance is parameterized by a formula $\Phi_n$ that consists of $m = m(n)$ 3XOR clauses, the set of which denoted by $C$, on $n$ boolean variables (i.e., each clause is $x_i x_j x_k = a_{ijk}$ for some combination of $(i,j,k)$ and $x_i, x_j, x_k \in \{\pm 1\}$). Thus, the objective function is $\Phi_n(x) = \frac{1}{m} \sum_{(i,j,k) \in C} \frac{1 + a_{ijk} x_i x_j x_k}{2}, x \in \{\pm 1\}^n$.

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. (Note that we have a slightly different formulation from [Bar14] that is slightly stronger but guaranteed by [Gri01].)

**Proposition 3.22 (Theorem 3.1 of [Bar14], due to Grigoriev)** For any $\epsilon > 0$, for every $n$ there exists a 3XOR instance $\Phi_n$ with $n$ variables and $m = O(n/\epsilon^2)$ clauses, such that $\text{OPT}(\Phi_n) \leq \frac{1}{2} + \epsilon$, but there exists a degree-$\Omega(n)$ perfectly satisfying pseudo-solution $\tilde{E}$.

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

In our framework, this implies a degree-$\Omega(n)$ value-(1, $1/2 + \epsilon$) integrality gap for the 3XOR problem.

4 Lower bounds on $h_{\text{sep}}$ and its applications

In this section, we will explain how the lower bounds on $h_{\text{sep}}$ are derived through reductions in the framework introduced in Section 3. To that end, we will illustrate the high level reduction path here and explain each reduction in details in following subsections.

$$3\text{XOR} \xrightarrow{R_1} 2\text{-out-of-4-SAT-EQ} \xrightarrow{R_2} \text{QMA}(2)\text{-Acc Prob} \xrightarrow{R_3} h_{\text{sep}}$$

There are three reductions $R_1, R_2, R_3$ respectively in the reduction path from 3XOR to $h_{\text{sep}}$. The start point is 3XOR as we introduced in Section 3.4. The two intermediate problems are roughly as follows:

- 2-OUT-OF-4-SAT-EQ: is a boolean polynomial optimization in which each instance is parameterized by a formula $\Phi_n$ that consists of 2-OUT-OF-4 clauses and EQ clauses.
Each 2-out-of-4 clause affects on 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 affects on 2 boolean variable \( x_i, x_j \in \{ \pm 1 \} \). The clause is satisfied if and only if \( x_i = x_j \).

- **QMA(2)-Acc Prob**: is a boolean polynomial optimization in which each instance refers to the acceptance probability of a specific QMA(2) protocol on some quantum witness generated by the boolean variables. The optimum value hence refers to the maximum acceptance probability of the protocol over all possible quantum witnesses generable in this way. The specific QMA(2) protocol aims to solve instances of the 2-out-of-4-SAT-EQ problem and has perfect completeness if the instance is satisfiable. Otherwise, the protocol accepts with probability bounded away from 1 by a constant (soundness).

We note that this chain of reductions is implicit in the earlier works of Aaronson et. al. [ABD+08] and Harrow and Montanaro [HM13], which show reductions from 3XOR to \( h_{\text{Sep}} \). However, our argument requires explicit analysis of the intermediate steps of the chain. One reason is that we must show that the individual reductions are pseudo-solution preserving (see definition 3.12). Another reason is to overcome a technical obstacle of the LRS theorem: it only applies to optimization problems over the boolean hypercube, and so cannot be directly applied to \( h_{\text{Sep}} \). In our proof, the problem QMA(2)-Acc Prob functions as a “restriction” of \( h_{\text{Sep}} \) to the boolean cube; we use LRS that QMA(2)-Acc Prob is hard to approximate, and thus so is \( h_{\text{Sep}} \). This step also crucially uses soundness of the specific QMA(2) protocol.

Precise definitions of each problem will appear in each corresponding subsection. All three reductions will be elaborated on in Section 4.1, 4.2, 4.3 respectively. We will briefly describe extensions to other ETH-based hardness results in Section 4.4. Finally we will discuss applications of this lower bound to the 2-to-4 norm and the dis-entangler conjecture in Sections 4.5 and 4.6.

### 4.1 From 3XOR to 2-out-of-4-SAT-EQ

We start with precise definition of the 2-out-of-4-SAT-EQ problem.

**Definition 4.1** 2-out-of-4-SAT-EQ is a boolean polynomial optimization problem with the following properties:

- **Instances**: for any \( n \), an instance is parameterized by a formula \( \Phi_n \) that consists of \( m(n) \) clauses, each of which is an instance of the predicate 2-out-of-4 or the predicate EQ, defined below.

- **Clauses**: The predicate 2-out-of-4(\( x_i, x_j, x_k, x_l \)) evaluates to 1 when exactly two of the inputs are equal to 1, and evaluates to 0 otherwise. It is easy to see that this can be expressed as 2-out-of-4(\( x, y, z, w \)) = \( q((x+y+z+w)) \), where \( q \) is a univariate degree-4 polynomial. The predicate EQ(\( x_i, x_j \)) evaluates to 1 when \( x_i = x_j \) and 0 otherwise; it can also be represented as a polynomial EQ(\( x_i, x_j \)) = 1 - \( \frac{1}{4} (x_i - x_j)^2 \).

- The objective function \( \Phi_n(x) \) evaluated at a boolean string \( x \in \{ \pm 1 \}^n \) counts the fraction of clauses in \( \Phi_n \) that are satisfied by \( x \). By the remarks above, it is clear that \( \Phi_n(x) \) be written as a polynomial of degree 4.
The problem is called 2-out-of-4-SAT if there is no EQ clause.

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 4.2** For all $m$, $n$, there exists a reduction that maps a given 3XOR instance $\Psi$ with $m$ clauses and $n$ variables, onto a 2-out-of-4-SAT-EQ instance $\Phi$ satisfying the following properties:

1. Every variable in $\Phi$ appears in at most $c$ clauses for some universal constant $c$.
2. $\Phi$ has $O(n + m)$ variables and $O(m)$ clauses.
3. If $\Psi$ is perfectly satisfiable, then so is $\Phi$.
4. If at most $1 - \delta$ fraction of the clauses of $\Psi$ are satisfiable, then at most $1 - \Omega(\delta)$ fraction of the clauses of $\Phi$ 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 [2]-[4]. Next, we achieve property [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 $y_{ijk}^1, y_{ijk}^2, y_{ijk}^3$, and generate the following three 2-out-of-4 clauses: 2-out-of-4($x_i, y_{ijk}^1, y_{ijk}^2, z$), 2-out-of-4($x_j, y_{ijk}^1, y_{ijk}^3, z$), and 2-out-of-4($x_k, y_{ijk}^2, y_{ijk}^3, z$). 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 $y_{ijk}^1, y_{ijk}^2, y_{ijk}^3$ that satisfies all three clauses; otherwise, at most two of the three clauses are satisfied for all assignments to $y_{ijk}^1, y_{ijk}^2, y_{ijk}^3$. In particular if $(x_i, x_j, x_k) = (1, 1, 1)$ then we set $(y_{ijk}^1, y_{ijk}^2, y_{ijk}^3) = (-1, -1, -1)$ and if $(x_i, x_j, x_k)$ has Hamming weight 1 then we set $(y_{ijk}^1, y_{ijk}^2, y_{ijk}^3) = (x_i, x_j, x_k)$. On the other hand, since each $y_{ijk}^c$ appears in two clauses, multiplying all the clauses $(x_i y_{ijk}^1 y_{ijk}^2 z)(x_j y_{ijk}^1 y_{ijk}^3 z)(x_k y_{ijk}^1 y_{ijk}^2 z) = x_i x_j x_k z$. 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 4.3 that fixes this issue while keeping all other properties.

Lemma 4.3 (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 $\text{OPT}(G) = 1$, then $\text{OPT}(G') = 1$.
4. If $\text{OPT}(G) = 1 - \epsilon$, then $\text{OPT}(G') \leq 1 - \eta \epsilon$ for constant $\eta$.

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 4.4 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 $\Phi_n$, we have for any polynomial $p(x)$ of degree at most $d - 4$,

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

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

Proof. We start with the degree-$\Omega(n)$ value-$(1, 1 + \epsilon)$ integrality gap from Proposition 3.22 for 3XOR. Using the reduction in Proposition 4.2, 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 4.2:

- 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 4.3).

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 3.22, 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)-Acc Prob

We start with description of the QMA(2) protocol and the formal definition of QMA(2)-Acc Prob. It is conceivable that we hope this QMA(2) protocol can solve 2-out-of-4-SAT-EQ problem thus we can reduce 2-out-of-4-SAT-EQ to it. Later on, we will use the direct connection between QMA(2) protocol and $h_{sep}$ to reduce the problem to $h_{sep}$. To that end, we employ a slight modification of the QMA(k) protocol due to Aaronson et. al. \cite{ABD+08} and turn the QMA(k) protocol into a QMA(2) by the Harrow-Montanaro protocol \cite{HM13}. Precisely,

**Proposition 4.5** 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 $\Phi$ of 2-out-of-4-SAT-EQ has $OPT(\Phi) = 1$, then protocol $P$ accepts with probability 1 with the following quantum witness $|\psi_x\rangle \otimes |\psi_x\rangle$, where

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

2. If any instance $\Phi$ of 2-out-of-4-SAT-EQ has $OPT(\Phi) \leq 1 - \delta$, then protocol $P$ accepts with probability at most $\epsilon$ 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 Aaronson et. al. \cite{ABD+08} and the Harrow-Montanaro protocol \cite{HM13}.

The QMA(k) protocol due to Aaronson et. al. \cite{ABD+08} can be used to solve 2-out-of-4 when $k = O(\sqrt{n})$. So we just need to show how to modify the protocol to handle equality clauses, and clauses with negations. (1) For clauses with negations, recall that to measure a clause 2-out-of-4($x_i, x_j, x_k, x_\ell$), Aaronson et. al. project to the subspace spanned by $\{|i\rangle, \cdots, |\ell\rangle\}$, and then check whether the state is orthogonal to $\frac{1}{2}(|i\rangle + |j\rangle + |k\rangle + |\ell\rangle)$. So if we want to check the clause 2-out-of-4($x_i, \neg x_j, x_k, x_\ell$), we project as before and check that the state is orthogonal to $\frac{1}{2}(|i\rangle - |j\rangle + |k\rangle + |\ell\rangle)$. (2) For an equality clause on $x_i$ and $x_j$, project to the subspace spanned by $\{|i\rangle, |j\rangle\}$, then check if the state is orthogonal to $|i\rangle - |j\rangle$. It is not hard to see the analysis therein works with this slight change.

Then we can apply the generic transition from any QMA(k) to QMA(2) in \cite{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 $\sqrt{n}\text{polylog}(n)$ qubits. These are grouped into registers of $\log(n)$ qubits each. The protocol consists of the following four tests, which we describe briefly (for full descriptions, see the original works):

1. **Product test:** This is due to \cite{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 where $m = \sqrt{n}\text{polylog}(n)$. In this test the verifier projects $A_iB_i$ onto the symmetric subspace for each $i$ and rejects if any $A_iB_i$ is found in the antisymmetric subspace.

2. **Symmetry test:** This is due to \cite{ABD+08}. In this test, the verifier projects $A_1, \ldots, A_m$ onto the symmetric subspace and similarly for $B_1, \ldots, B_m$.

---

\textsuperscript{3}We use the convention that $\tilde{O}(\cdot)$ hides constants as well as polylog($n$) terms.
3. Uniformity test: Also due to [ABD+08], q.v. for details.

4. Satisfiability test: Also due to [ABD+08], q.v. for details. We modify this test slightly as described above.

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 $\ell$ times in parallel, the new soundness is at most $s^\ell$.

We are ready to formally define QMA(2)-Acc Prob. Let protocol $P$ be from Proposition 4.5.

**Definition 4.6** QMA(2)-Acc Prob is a boolean polynomial optimization problem with the following restriction:

- **Instances**: for any $n$, let $P_n$ denote the QMA(2) protocol from Proposition 4.5 that solves instances of 2-out-of-4-SAT-EQ on $n$ boolean variables. The instance $\Psi$ of QMA(2)-Acc Prob is parameterized by instance $\Phi$ of 2-out-of-4-SAT-EQ. The objective function $\Psi : \{\pm 1\}^n \to [0,1]$ is given by

$$\Psi(x) = \Pr[P_n \text{ accepts on input } \Phi \text{ and witness } |\psi_x\rangle \otimes |\psi_x\rangle],$$

where $x \in \{\pm 1\}^n$, $M^\Phi_n$ is the POVM corresponding to the acceptance in protocol $P$ on input $\Phi$ and $|\psi_x\rangle$ is defined in (4.1). It is not hard to verify that $\Psi(x)$ is indeed a degree-$\tilde{O}(\sqrt{n})$ polynomial, although its explicit form is irrelevant for our purpose.

**4.2.1 Integrality gap for QMA(2)-Acc Prob**

We are ready to establish the integrality gap for QMA(2)-Acc Prob 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)-Acc Prob as in the Definition 4.6.

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

**Proof.** For any $n$, let $\Psi_n \in$ 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 4.4 and $\mu_n$ be the corresponding pseudo-solution ($\delta$ is the constant therein). Let $\Psi_n$ be its reduced instance of QMA(2)-Acc Prob. For any constant $\epsilon > 0$, by Property (2) of Proposition 4.5 and the fact $\text{OPT}(\Psi_n) \leq 1 - \delta$, we have $\text{OPT}(\Psi_n) \leq \epsilon$. Then it suffices to establish a $(\Omega(n), 0)$ pseudo-solution for $\Psi_n$.

To that end, we claim that $\mu_n$ is also a $(\Omega(n), 0)$ pseudo-solution for $\Psi_n$. Note that the feasible sets for $\Phi_n$ and $\Psi_n$ are the same. It suffices to show $\mathbb{E}_{x \sim \mu}[\Psi_n(x)] = 1$. Observe that by linearity of $\mathbb{E}[]$ and $\text{tr}(\cdot)$, we have

$$\mathbb{E}_{x \sim \mu}[\Psi_n(x)] = \mathbb{E}_{x \sim \mu}[\text{tr}(M^\Psi_n |\psi_x\rangle \langle \psi_x| \otimes |\psi_x\rangle \langle \psi_x|)] = \text{tr}(M^\Psi_n \mathbb{E}_{x \sim \mu}[|\psi_x\rangle \langle \psi_x| \otimes |\psi_x\rangle \langle \psi_x|]).$$
Thus, we define \( \tilde{\rho} \) as
\[
\tilde{\rho} = \mathbb{E}_{x \sim \mu} [ |\psi_x\rangle \langle \psi_x| \otimes |\psi_x\rangle \langle \psi_x|].
\] (4.2)

Now, we need to calculate the expectation value of \( M^{\Psi_n} \), the POVM element corresponding to the “yes” outcome of the protocol. Recall from proposition 4.5 that our protocol is obtained by parallel repetition of the protocol of [HM13], where the number of repetitions is constant. Thus, \( M^{\Psi_n} \) 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 at least of the terms corresponds to a “no” outcome of a test. To show that the state \( \tilde{\rho} \) passes \( M^{\Psi_n} \) 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. Tr\[ (M^{\text{no}}, \text{symmetry test} \otimes M^{\text{rest}}) \] = 0.

2. **Uniformity test:** Recall that in the uniformity test, Arthur chooses a matching \( M \) on \([n]\), and the measures each subsystem in an orthonormal basis containing
\[
|\pm\rangle_{ij} = \frac{1}{\sqrt{2}} (|i\rangle \pm |j\rangle)
\]
for every \((i,j) \in M\). 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
\[
\Pr[\text{Uniformity test failure}] = \Pr[\tilde{\rho} |+\rangle_{ij} \langle +|_{ij} \otimes |--\rangle_{ij} \langle --|_{ij} \otimes M^{\text{rest}}]
\]
\[
= \mathbb{E}_{x \sim \mu}[(x_i + x_j)^2(x_i - x_j)^2q(x)], \text{ for some polynomial } q(x)
\]
\[
= \mathbb{E}_{x \sim \mu}[(x_i^2 + x_j^2 + 2x_ix_j)(x_i^2 + x_j^2 - 2x_ix_j)q(x)]
\]
\[
= \mathbb{E}_{x \sim \mu}[(2 + 2x_ix_j)(2 - 2x_ix_j)q(x)] = \mathbb{E}_{x \sim \mu}[4(1 - x_i^2 x_j^2)q(x)] = 0.
\]
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 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

\[ \Pr[\text{Success}] = \text{Tr} [(I - \Pi_{C_1}) \otimes (I - \Pi_{C_2}) \otimes \cdots \otimes (I - \Pi_{C_m}) \otimes M^\text{rest} \tilde{\rho}] \]

\[ \propto \tilde{\mathbb{E}}_{x \sim \mu} \left[ (1 - \frac{1}{4} \left( \sum_{x \in C_1} x \right)^2) \cdots (1 - \frac{1}{4} \left( \sum_{x \in C_m} x \right)^2) \cdots \right]. \]

Now, we know that the pseudo-solution \( \mu \) 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 \( \Pr[\text{Success}] = 1 \) 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 \( \Pr[\text{Success}] = 1 \) 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.

### 4.3 Lower bounds for DPS and general SDPs

In this section, we start with the last piece of the reduction, i.e, \( R_3 \) from QMA(2)-Acc Prob to \( h_{\text{Sep}} \). The reduction is straightforward given their definitions. To be precise, for any instance \( \Psi \) of QMA(2)-Acc Prob that is parameterized by instance \( \Phi \) from 2-out-of-4-SAT-EQ on \( n \) variables, we define the following \( h_{\text{sep}}(d, d) \) problem: the dimension \( d \) is the same as \( |\psi_x\rangle \) (i.e., \( d = 2^{\tilde{O}(\sqrt{n})} \)) and the objective function is the inner product of \( M^\Phi \) and the separable states, where \( |\psi_x\rangle \) and \( M^\Phi \) are from Definition 4.6. Thus, we have

**Proposition 4.8** For any \( 0 < \delta, \epsilon < 1 \), define QMA(2)-Acc Prob according to Proposition 4.5 and the reduction \( R_3 \) accordingly. Then \( R_3 \) is an \( (O(\epsilon), \epsilon) \)-approximate embedded reduction.

**Proof.** We prove the state by verifying each property according to definitions.

- \( R_3 \) is \( (O(\epsilon), \epsilon) \)-approximate—for the case of 2-out-of-4-SAT (without equality clauses), this follows from the soundness of the protocol established in Corollary 12 of [HM13]. More specifically, if the optimal value of QMA(2)-Acc Prob is at most \( \epsilon \), then the true value of the underlying 2-out-of-4-SAT instance \( \Phi \) is at most \( O(\epsilon) \), so by the soundness of the protocol, \( h_{\text{sep}}(M^\Phi) \leq O(\epsilon) \). It is easy to see that the soundness is maintained even with the addition of equality clauses.

- \( R_3 \) is embedded because for each \( x \in \{\pm 1\}^n \), it is mapped to \( |\psi_x\rangle \otimes |\psi_x\rangle \) where \( |\psi_x\rangle \) is from [4.1]. Moreover, for any instance of QMA(2)-Acc Prob that is parameterized by \( \Phi \) from 2-out-of-4-SAT-EQ, the value \( \Psi(x) \) is exactly the inner product of \( M^\Phi \) and \( |\psi\rangle \otimes |\psi\rangle \).
Now we can state our main SDP lower bound for $h_{\text{Sep}}$ due to Proposition 3.19, 3.20.

**Theorem 4.9** For any $0 < \epsilon, \eta$, $\epsilon + \eta < 1$, any SDP relaxation $A$ of $h_{\text{Sep}(d,d)}$ achieving a $(1 - \eta, \epsilon)$ approximation must have size at least

$$r(d) \geq \Omega\left(d\log(d)/\text{polylog log}(d)\right).$$

**Proof.** By Proposition 3.19, composition of the reduction $R_3$ and the one from $h_{\text{Sep}}$ to the SDP relaxation $A$ gives a SDP relaxation of the QMA(2)-Acc Prob problem. Given the parameters in Proposition 4.8, we have that $A$ is a $(1 - \eta, O(\epsilon))$-approximate SDP relaxation of QMA(2)-Acc Prob. By Proposition 3.20 and Proposition 4.7, we have the size $r(n)$ of any $(1 - \eta, O(\epsilon))$-approximate $A$ (in fact, any $(1, O(\epsilon))$-approximate $A$) is at least,

$$r(n) \geq \Omega\left(\left(\frac{n}{\log n}\right)^{\Omega(n)}\right) \geq \Omega\left(n^{\Omega(n)}\right).$$

Given $d = 2^{O(\sqrt{n}\text{polylog}(n))}$, we have,

$$r(d) \geq \Omega(\exp(\Omega(n)\log(n))) \geq \Omega\left(d^{\Omega(\sqrt{n})\log(n)/\text{polylog}(n)}\right) \geq \Omega\left(d^{\log(d)/\text{polylog log}(d)}\right).$$

**DPS lower bound**

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

**Corollary 4.10** For any constant $\epsilon$, 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 $k$-level of DPS, i.e., DPS$(M) = 1$ for $k \leq o(\log d/\text{polylog log} d)$.

**Proof.** We take $M_d$ to be $M^\Psi$ from the QMA(2) protocol, and $\tilde{\rho}$ from (4.2). The state $\tilde{\rho}$ arises as the reduced density matrix of a fully symmetric state $\rho$ on $\Omega(n)$ registers and is thus a $O(\sqrt{n})$-extendible state. The statement follows because $d = 2^{O(\sqrt{n})}$.

**4.4 Other hardness results**

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 theorem 4.9, 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]. This protocol solves an NP-hard graph coloring problem in QMA(2) with completeness 1 and soundness $1 - \tilde{O}(1/n)$. Schematically, the proof of the hardness result is:

$$3\text{XOR}(n) \implies \text{Graph-3-coloring}(n) \implies \text{QMA(2)-Acc Prob} \implies h_{\text{Sep}(n,n)},$$

where Graph-3-coloring$(n)$ is the problem of deciding whether a graph of $n$ vertices is 3-colorable, and QMA(2)-Acc Prob is defined as before but with reference to the honest witnesses of the protocol of [LNN12]. To achieve hardness for $h_{\text{Sep}}$, 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, and it is straightforward to verify that it satisfies the hypothesis of proposition 3.16 with a constant value of $\kappa$; 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 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. This gives us a degree $\Omega(n)$ value 1 pseudo-solution for QMA(2)-Acc Prob. Finally, the reduction in the last arrow is a $(1 - \tilde{O}(1/n), 1 - \eta)$-approximate embedding reduction for some constant $\eta$, by the soundness of the protocol. When we chain these reductions together, and apply the LRS theorem, we obtain that any SDP relaxation achieving a $(1, 1 - \tilde{O}(1/n))$ approximation to $h_{\text{Sep}(n,n)}$ must have size at least $\exp(\Omega(n))$.

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 completeness $1 - \exp(-\Omega(\sqrt{n}))$ 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 $h_{\text{Sep}^{O(\sqrt{n})}(n)}(M)$, where Sep$^k(n)$ means $k$-partite separable states, and $M$ is restricted to be a bell measurement. The schematic diagram for this case is

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

The arguments are very similar to those in the previous result; when we work out the parameters, we obtain that any SDP relaxation that achieves a $(1 - \exp(-\Omega(\sqrt{n})), 1 - \eta)$ approximation to $h_{\text{Sep}^{O(\sqrt{n})}(n)}(M)$ for general Bell measurements $M$ and the appropriate constant $\eta$ must have size at least $\exp(\Omega(n))$.

### 4.5 Application to $2 \to 4$ norm

If $A \in \mathbb{R}^{m \times n}$ then define the $2 \to 4$ norm of $A$ to be

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

In [BBH+12] it was shown that computing the $2 \to 4$ norm was a special case of computing $h_{\text{Sep}}$ and that in turn there was an approximation-preserving reduction from $h_{\text{Sep}}$ 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 4.11 If $S$ is an SDP relaxation of $\|A\|_{2\rightarrow 4}$ achieving multiplicative error of $C = O(1)$ then it must have size at least 
\[ r(d) \geq \Omega(d^{\log(d)/\polylog\log(d)}). \]

4.6 The no-disentangler conjecture

One application of our result is to prove a version of the Approximate Disentangler Conjecture, 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 4.12 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 $\Lambda : D(\mathcal{H}) \rightarrow D(\mathcal{K} \otimes \mathcal{K})$ is an $(\epsilon, \delta)$-approximate disentangler if

- For every $\rho \in D(\mathcal{H})$, $\Lambda(\rho)$ is $\epsilon$-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 $\delta$-close in trace distance to $\sigma$.

Our result is the following:

Theorem 4.13 Suppose $\Lambda : D(\mathcal{H}) \rightarrow D(\mathcal{K} \otimes \mathcal{K})$ is an $(\epsilon, \delta)$-approximate disentangler with $\epsilon + \delta < 1$, and let $d = \dim(\mathcal{K})$. Then
\[ \dim(\mathcal{H}) \geq \Omega(d^{\log(d) / \polylog\log(d)}), \]

This is weaker than Watrous’s original formulation, which had $\dim(\mathcal{H}) \geq \exp(d)$. However, it matches the conditional result obtained from ETH.

Proof. We show that a disentangler can be used as an SDP relaxation to $h_{\text{Sep}}$, thus allowing us to apply theorem 4.9. 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 $R_\Lambda$ from $h_{\text{Sep}}$ to $h_\Lambda(M)$, that achieves a $(s + \epsilon, s)$ approximation. This reduction simply maps an instance $M$ of $h_{\text{Sep}}$ to the instance $h_\Lambda(M)$ given by the same measurement operator $M$. The embedding property follows from the definitions: for every separable $\sigma$, there exists a $\rho$ 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 $\rho$, $\Lambda(\rho)$ is $\epsilon$-close to some separable $\sigma$. 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 semidefinite relaxation of $h_{\text{Sep}}$. So, by applying theorem 4.9 we conclude that $\dim(\mathcal{H}) \geq d^{\frac{\log(d) / \polylog\log(d)}{}}$.

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 rom $h_{\text{Sep}}$ to $h_{\text{Disentangled}}$. We will fix this by using the following gadget: Let $B_\delta$ be the set of states in $D(\mathcal{K} \otimes \mathcal{K})$ of trace norm less than or equal to $\delta$. Then, given an $(\epsilon, \delta)$-disentangler $\Lambda$, we define a new map $\tilde{\Lambda} : D(\mathcal{H}) \oplus B_\delta \rightarrow D(\mathcal{K} \otimes \mathcal{K})$ by
\[ \tilde{\Lambda}(\rho \oplus \sigma) \equiv \Lambda(\rho) + \sigma. \]
We claim that for every separable $\tau \in \text{Sep}(K, K)$, there exists a preimage $\rho \in D(H), \sigma \in B_\delta$ with $\Lambda(\rho \oplus \sigma) = \tau$. Indeed, since $\Lambda$ is an $(\epsilon, \delta)$-disentangler, we know that $\tau$ had an approximate preimage $\rho$ 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 $\rho \in D(H), \sigma \in B_\delta$, $\tilde{\Lambda}(\rho \oplus \sigma)$ is within $\epsilon + \delta$ in trace distance of some separable state. To see this, note that $\Lambda(\rho)$ is within $\epsilon$ distance of some separable state $\tau$, and since $\|\sigma\|_1 \leq \delta$, adding $\sigma$ can increase the distance to $\tau$ by at most $\sigma$.

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

$$\max_{\rho, \sigma^+, \sigma^-} \text{Tr}[M \tilde{\Lambda}(\rho \oplus \sigma)]$$

such that

$$\text{Tr}[\rho] = 1$$

$$\text{Tr}[\sigma^+ + \sigma^-] \leq \delta$$

$$\rho, \sigma^+, \sigma^- \succeq 0.$$

This SDP implements the constraint $\|\sigma\|_1 \leq \delta$. As before, consider the reduction from $h_{\text{Sep}}$ 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 relaxation of $h_{\text{Sep}}$. So once again, applying theorem 4.9 tells us that the dimension of the SDP $h_{\tilde{\Lambda}}$ must be at least $\Omega(d \log^2 d / \text{polylog} \log d)$. Now, the dimension of $h_{\tilde{\Lambda}}$ is equal to $\dim(H) + d$, so all together we get

$$\dim(H) \geq \Omega(d \log^2 d / \text{polylog} \log d) - d \geq \Omega(d \log^2 d / \text{polylog} \log d).$$

5 Lower Bounds for Entangled Games

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

**Definition 5.1** 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 5.2** A (deterministic) classical strategy for a nonlocal game consists of functions $f_1, f_2 : Q \rightarrow A$, with $f_1(q)$ being the answer that player 1 gives to question $q$, and likewise for $f_2(q)$. The classical value $\omega_{\text{classical}}(G)$ of a game $G$ is the maximum winning probability of a classical strategy for $G$. 

28
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 5.3** 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 $\{A^a_q \otimes I_2\}_{a \in A}$ acting nontrivially on only player 1’s Hilbert space, and
- for every question $q \in Q$, a POVM $\{I_1 \otimes B^a_q\}_{a \in A}$ 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 relaxations, as well as an explicit counterexample 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].

$$3\text{XOR} \Rightarrow \omega_{\text{honest classical}} \Rightarrow \omega_{\text{entangled}}$$

The intermediate problems are

- 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 5.4 (Lemma 8 of [IKM09])** Let $\Phi$ be a 3-CSP over $n$ variables. Then there exists a 2-player quantum game $G_\Phi$ such that for some constant $c$,

$$\text{MAX-SAT}(\Phi) \leq \omega_{\text{classical}}(G_\Phi) \leq 1 - \frac{1 - \text{MAX-SAT}(\Phi)}{3},$$

$$\omega_{\text{entangled}}(G_\Phi) \leq 1 - \frac{c(1 - \text{MAX-SAT}(\Phi))^2}{n^2}.$$
Definition 5.5 Let $\Phi$ be a 3-CSP. Then the oracularization of $\Phi$ is a 2-player entangled game $G_{\Phi}$, where each player receives three indices $i_1, i_2, i_3$ corresponding to a random clause in $\Phi$, and returns three answer bits $a_1, a_2, a_3$. With equal probability, the verifier performs one of the following checks:

1. **Simulation check:** The verifier chooses a player at random and verifies that the answers $a_1a_2a_3$ satisfy the clause associated with variables $i_1i_2i_3$.

2. **Consistency check:** If the two players are both asked about the same variable, the verifier checks that their answers for this variable agree.

In our application of this result, we will take the 3-CSP to be $3\text{XOR}$. We say that the players are playing honestly according to assignment $x$ if each player responds to the question $i_1i_2i_3$ with the answers $x_{i_1}x_{i_2}x_{i_3}$. For a given $3\text{XOR}$ instance $\Phi$ and assignment $x$, let $f_{\Phi}(x) : \{0, 1\}^n \to \{0, 1\}$ be the winning probability in the game $G_{\Phi}$ of the honest strategy according to $x$.

Definition 5.6 The problem $\omega_{\text{honest classical}}(G_{\Phi})$ is the optimization problem of maximizing $f_{\Phi}(x)$ over $x \in \{\pm 1\}^n$.

Lemma 5.7 Let $\Phi$ be a $3\text{XOR}$ instance produced by Proposition 3.22. Then there exists a degree-$\Omega(n)$ value-$(1 - \epsilon, 1 - c)$ integrality gap for $\omega_{\text{honest classical}}(G_{\Phi})$ for all $\epsilon$.

Proof. 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 $\tilde{\mathcal{E}}$ produced by proposition 3.22. 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 $3\text{XOR}$ 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

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

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

$$\tilde{\mathcal{E}}[V_{b, A}] = \frac{1}{2} + \frac{1}{2} b \tilde{\mathcal{E}}[x_i x_j x_k] = 1.$$

- **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 $\tilde{\mathcal{E}}$.

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

For a general 2-player game $G$, define the optimization problem $\omega_{\text{entangled}}(G)$ to be the optimization of the winning probability of game $G$ over all two-player entangled strategies.

**Theorem 5.8** Suppose $S_n$ is a sequence of SDP relaxations to the problem $\omega_{\text{entangled}}(G)$ of size $r_n$, achieving an $(1-\epsilon(n), 1-\delta(n))$-approximation, where $\delta(n) = O(1/n^2)$ and $\epsilon(n) < \delta(n)$. Then $r_n \geq \Omega \left( \left( \frac{n}{\log n} \right)^{\Omega(n)} \right)$.

**Proof.** As before, let $G_\Phi$ be the oracularized game associated with the 3XOR instance $\Phi$, 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 $\omega_{\text{entangled}}(G)$. Moreover, by lemma 5.4, this reduction is $(1-c/n^2, 1-c')$-approximate for constants $c, c'$. Thus, any SDP relaxation of size $r_n$ for $\omega_{\text{entangled}}(G)$ that achieves a $(1-\epsilon(n), 1-c/n^2)$ approximation implies an SDP relaxation of size $r_n$ for $\omega_{\text{Honest Classical}}(G)$ that achieves a $(1-\epsilon(n), 1-c')$ approximation. Now, by proposition 3.20, $r_n \geq \Omega \left( \left( \frac{n}{\log n} \right)^{d/4} \right)$, where $d$ is the degree of an integrality gap for an instance of $\omega_{\text{Honest Classical}}(G)$ of size $n/2$. By lemma 5.7, $d = \Omega(n)$. Substituting this into the expression above yields the desired result.

5.2 An explicit lower bound for ncSoS

In the previous section we gave a lower bound on the size of general SDP relaxations to this problem. We will now present an explicit lower bound on the well-known 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 $E_x \sim \mu f(x)$ over pseudodistributions $\mu$ that obey certain constraints. Likewise, in the ncSoS hierarchy, the winning probability $\omega$ 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 polynomial. A non-commuting pseudoexpectation satisfies conditions similar to an ordinary pseudoexpectation operator:

**Definition 5.9** An degree $d$ ncSoS pseudoexpectation is a linear map $E[\cdot]$ that maps polynomials in the provers’ measurement operators $\{A_q^a\}, \{B_q^b\}$ to real numbers. This map satisfies the following properties:

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

**Theorem 5.10** For every $n$ there exists a two-player entangled game $G$ with $O(n)$ questions and three-bit answers, such that $\omega_{\text{entangled}}(G) \leq 1 - c/n^2$ for some constant $c$, but there exists $m = \Omega(n)$ and a pseudoexpectation of degree $m$ according to which the game value is 1.
Proof. Start with a 3XOR instance with maximum satisfiable fraction $1/2 + \epsilon$. Then lemma 5.4 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 $\{A_{i_1i_2i_3}^{a_1a_2a_3}\}$, $\{B_{i_1i_2i_3}^{b_1b_2b_3}\}$, where $(i_1, i_2, i_3)$ are the indices of three variables in the assignment. 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 $A_{i_1i_2i_3}^{a_1a_2a_3} = B_{i_1i_2i_3}^{a_1a_2a_3} = C_{i_1}^{a_1}C_{i_2}^{a_2}C_{i_3}^{a_3}$ where the operators $\{C_{i_1}^{0}, C_{i_1}^{1}\}$ form a projective measurement for every index $i$. For convenience, we will henceforth work with the observables $C_{i_1}^{0} = C_{i_1}^{1} - C_{i_1}^{1}$; these square to identity. Now, let $\tilde{E}$ be the Grigoriev pseudoexpectation operator for the 3XOR instance. We define an ncSoS pseudoexpectation $\tilde{E}'$ as follows:

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

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 $x_ix_jx_k = b$, we want to verify that player $A$’s answers multiply together to $b$. For every clause $b = x_ix_jx_k$, we have a term

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

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

$$\tilde{E}'[V_{b,A}] = \sum_{a_i,a_j,a_k} (a_i a_j a_k b) \tilde{E}'[A_{x_i x_j x_k}^{a_i a_j a_k} \otimes I_B]$$

$$= \sum_{a_i,a_j,a_k} a_i a_j a_k b \tilde{E}'[C_{x_i}^{a_i}C_{x_j}^{a_j}C_{x_k}^{a_k} \otimes I_B]$$

$$= b \tilde{E}'[C_{x_i}C_{x_j}C_{x_k} \otimes I_B]$$

$$= b \tilde{E}[x_i x_j x_k]$$

$$= 1.$$
about the same bit.

\[ V_{i,A,B} = \mathbb{E}_{j,k,p,q} \sum_{a_1 a_2 a_3 b_1 b_2 b_3} (a_1 b_1) A_{x_i x_j x_k}^{a_1 a_2 a_3} \otimes B_{x_i x_p x_q}^{b_1 b_2 b_3} \]

\[ \tilde{\mathbb{E}}'[V_{i,A,B}] = \mathbb{E}_{j,k,p,q} \sum_{a_1 a_2 a_3 b_1 b_2 b_3} (a_1 b_1) \tilde{\mathbb{E}}'[A_{x_i x_j x_k}^{a_1 a_2 a_3} \otimes B_{x_i x_p x_q}^{b_1 b_2 b_3}] \]

\[ = \sum_{a_1 b_1} (a_1 b_1) \mathbb{E}'[C_{i}^{a_1} C_{i}^{b_1}] \]

\[ = \tilde{\mathbb{E}}'[C_i C_i] \]

\[ = \tilde{\mathbb{E}}[x_i^2] \]

\[ = 1. \]

\[ \tilde{\mathbb{E}}'[C_i C_i] = \tilde{\mathbb{E}}[x_i^2] = 1. \]

**Acknowledgement**

AWH was funded by NSF grant CCF-1111382 and 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.

**References**


John William Helton and Mihai Putinar. Positive polynomials in scalar and matrix variables, the spectral theorem and optimization, 2006.


A Proofs omitted from the main text

A.1 Proof of Proposition 3.20

Proof. Let us recall some key concepts from [LRS14].

Definition A.1 Let \( f_\Phi : \{\pm 1\}^n \to [0,1] \) be a set of functions on the boolean cube parametrized by \( \Phi \), and let \( 0 \leq s < c \leq 1 \). Then the matrix \( M_{c,s}^{n,f} \) is defined by

\[
M_{c,s}^{n,f} : \{ \Phi : \max_{x \in \{0,1\}^n} f_\Phi(x) \leq s \} \times \{0,1\}^n \to \mathbb{R}, \quad M(\Phi,x) = c - f_\Phi(x).
\]

The main result of [LRS14] relates the size of any SDP relaxation with the psd rank of \( M_{c,s}^{n,\Phi} \).

Proposition A.2 (Proposition 6.1 of [LRS14]) There exists a \((c(n),s(n))\)-approximate SDP relaxation with size \( r(n) \) if and only if \( \text{rk}_{psd}(M_{c(n),s(n)}^{n,\Phi}) \leq r(n) \), for any large \( n \).

To lower bound the psd rank of \( M_{c,s}^{n,\Phi} \), we turn to the following pattern matrix:

Definition A.3 Let \( g : \{\pm 1\}^m \to [0,1] \) be a function on the boolean cube, and let \( n \geq m \). For a subset \( S \subset [n] \), \( S = \{i_1, \ldots, i_{|S|}\} \) and a vector \( x \in \mathbb{R}^n \), we let \( x_S \equiv (x_{i_1}, x_{i_2}, \ldots, x_{i_{|S|}}) \). Then the matrix \( M_{n,\Phi}^g \) is defined by

\[
M_{n,\Phi}^g : \{ S \subseteq [n] : |S| = m \} \times \{0,1\}^n \to \mathbb{R}, \quad M_{n,\Phi}^g(S,x) \equiv g(x_S).
\]

Let \( m = \lfloor n/2 \rfloor \) and \( g(x) = c - f_\Phi(x) \) for \( x \in \{\pm 1\}^m \). Then it is easy to see that \( M_{n,\Phi}^g \) is a sub-matrix of \( M_{c,s}^{n,\Phi} \). We then make use of the following lower bound for \( \text{rk}_{psd}(M_{n,\Phi}^g) \) as a lower bound for \( \text{rk}_{psd}(M_{c,s}^{n,\Phi}) \).

Proposition A.4 (Theorem 1.8 of [LRS14]) Let \( m \geq 1 \) and \( g : \{\pm 1\}^m \to [0,1] \). For \( n \geq 2m \), if \( d + 2 = \deg_{SOS}(g) \), then

\[
1 + n^{1+d/2} \geq \text{rk}_{psd}(M_{n,\Phi}^g) \geq C \left( \frac{n}{\log n} \right)^{d/4}.
\]

As a consequence, we have

\[
\text{rk}_{psd}(M_{c(n),s(n)}^{n,\Phi}) \geq \text{rk}_{psd}(M_{n,\Phi}^g) \geq C \left( \frac{n}{\log n} \right)^{d/4},
\]

where \( d = \deg_{SOS}(g) \) and \( g = c - f_\Phi(x) \). Because the existence of a degree-\(d(n)\) value \((1 - \epsilon(n), s(n))\) integrality gap, by Proposition 3.5 we choose \( c = 1 - \epsilon(m) \) and have \( d = \deg_{SOS}(g) - 2 \geq d(m) - 2 \). Thus, we have any \((1 - \epsilon(m), s(n))\)-approximate SDP relaxation must have its size \( r(n) \) lower bounded by

\[
r(n) \geq C \left( \frac{n}{\log n} \right)^{(d(m)-2)/4} = C \left( \frac{n}{\log n} \right)^{\Omega(d(|n/2|))}.
\]