

## MIT Open Access Articles

*Hypercontractivity, sum-of-squares proofs, and their applications*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Boaz Barak, Fernando G.S.L. Brandao, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. 2012. Hypercontractivity, sum-of-squares proofs, and their applications. In Proceedings of the 44th symposium on Theory of Computing (STOC '12). ACM, New York, NY, USA, 307-326.

**As Published:** <http://dx.doi.org/10.1145/2213977.2214006>

**Publisher:** Association for Computing Machinery (ACM)

**Persistent URL:** <http://hdl.handle.net/1721.1/80386>

**Version:** Original manuscript: author's manuscript prior to formal peer review

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike 3.0



# Hypercontractivity, Sum-of-Squares Proofs, and their Applications

Boaz Barak\*      Fernando G.S.L. Brandão†      Aram W. Harrow‡  
 Jonathan Kelner§      David Steurer¶      Yuan Zhou||

October 31, 2012

## Abstract

We study the computational complexity of approximating the  $2$ -to- $q$  norm of linear operators (defined as  $\|A\|_{2 \rightarrow q} = \max_{v \neq 0} \|Av\|_q / \|v\|_2$ ) for  $q > 2$ , as well as connections between this question and issues arising in quantum information theory and the study of Khot’s Unique Games Conjecture (UGC). We show the following:

1. For any constant even integer  $q \geq 4$ , a graph  $G$  is a *small-set expander* if and only if the projector into the span of the top eigenvectors of  $G$ ’s adjacency matrix has bounded  $2 \rightarrow q$  norm. As a corollary, a good approximation to the  $2 \rightarrow q$  norm will refute the *Small-Set Expansion Conjecture* — a close variant of the UGC. We also show that such a good approximation can be obtained in  $\exp(n^{2/q})$  time, thus obtaining a different proof of the known subexponential algorithm for SMALL-SET EXPANSION.
2. Constant rounds of the “Sum of Squares” semidefinite programming hierarchy certify an upper bound on the  $2 \rightarrow 4$  norm of the projector to low-degree polynomials over the Boolean cube, as well certify the unsatisfiability of the “noisy cube” and “short code” based instances of UNIQUE GAMES considered by prior works. This improves on the previous upper bound of  $\exp(\log^{O(1)} n)$  rounds (for the “short code”), as well as separates the “Sum of Squares”/“Lasserre” hierarchy from weaker hierarchies that were known to require  $\omega(1)$  rounds.
3. We show reductions between computing the  $2 \rightarrow 4$  norm and computing the *injective tensor norm* of a tensor, a problem with connections to quantum information theory. Three corollaries are: **(i)** the  $2 \rightarrow 4$  norm is NP-hard to approximate to precision inverse-polynomial in the dimension, **(ii)** the  $2 \rightarrow 4$  norm does not have a good approximation (in the sense above) unless 3-SAT can be solved in time  $\exp(\sqrt{n} \text{ poly } \log(n))$ , and **(iii)** known algorithms for the quantum separability problem imply a non-trivial additive approximation for the  $2 \rightarrow 4$  norm.

---

\*Microsoft Research New England.

†Department of Physics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

‡Department of Computer Science & Engineering, University of Washington. Much of the work done while visiting Microsoft Research New England.

§Department of Mathematics, Massachusetts Institute of Technology.

¶Microsoft Research New England.

||Department of Computer Science, Carnegie Mellon University. Much of the work done while the author was an intern at Microsoft Research New England.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Our Results</b>	<b>1</b>
2.1	Algorithms . . . . .	2
2.1.1	Subexponential algorithm for “good” approximation . . . . .	2
2.1.2	Polynomial algorithm for specific instances . . . . .	2
2.1.3	Quasipolynomial algorithm for additive approximation . . . . .	3
2.2	Reductions . . . . .	4
2.2.1	Hypercontractivity and small set expansion . . . . .	4
2.2.2	Hypercontractivity and the injective tensor norm . . . . .	5
2.3	Relation to the Unique Games Conjecture . . . . .	5
<b>3</b>	<b>The SoS hierarchy</b>	<b>6</b>
3.1	Basic properties of pseudo-expectation . . . . .	8
3.2	Why is this SoS hierarchy useful? . . . . .	9
<b>4</b>	<b>Overview of proofs</b>	<b>10</b>
4.1	Subexponential algorithm for the 2-to-q norm . . . . .	10
4.2	Bounding the value of SoS relaxations . . . . .	10
4.3	The 2-to-4 norm and small-set expansion . . . . .	11
4.4	The 2-to-4 norm and the injective tensor norm . . . . .	11
4.5	Definitions and Notation . . . . .	12
<b>5</b>	<b>The Tensor-SDP algorithm</b>	<b>13</b>
<b>6</b>	<b>SoS succeeds on Unique Games integrality gaps</b>	<b>14</b>
6.1	Proof sketch of Theorem 6.1 . . . . .	14
6.2	Invariance Principle for Fourth Moment . . . . .	16
6.3	Interlude: Independent Rounding . . . . .	18
6.4	Dictatorship Test for Small-Set Expansion . . . . .	19
6.5	Dictatorship Test for Unique Games . . . . .	20
6.6	Influence Decoding . . . . .	20
6.7	Certifying Small-Set Expansion . . . . .	22
6.8	Putting Things Together . . . . .	22
6.9	Refuting Instances based on Short Code . . . . .	22
<b>7</b>	<b>Hypercontractivity of random operators</b>	<b>23</b>
<b>8</b>	<b>The 2-to-q norm and small-set expansion</b>	<b>26</b>
<b>9</b>	<b>Relating the 2-to-4 norm and the injective tensor norm</b>	<b>30</b>
9.1	Equivalent maximizations . . . . .	30
9.1.1	Injective tensor norm and separable states . . . . .	30
9.1.2	Connection to the 2-to-4 norm . . . . .	32
9.2	Hardness of approximation for the 2-to-4 norm . . . . .	33
9.2.1	Hardness of approximation for projectors . . . . .	37
9.3	Algorithmic applications of equivalent formulations . . . . .	38
9.3.1	Approximation guarantees and the proof of Theorem 2.3 . . . . .	40
9.3.2	Gap instances . . . . .	41
<b>10</b>	<b>Subexponential algorithm for the 2-to-q norm</b>	<b>42</b>
	<b>References</b>	<b>44</b>
<b>A</b>	<b>More facts about pseudo-expectation</b>	<b>48</b>

<b>B</b>	<b>Norm bound implies small-set expansion</b>	<b>49</b>
<b>C</b>	<b>Semidefinite Programming Hierarchies</b>	<b>49</b>
C.1	Example of Max Cut . . . . .	50

# 1 Introduction

For a function  $f: \Omega \rightarrow \mathbb{R}$  on a (finite) probability space  $\Omega$ , the  $p$ -norm is defined as  $\|f\|_p = (\mathbb{E}_\Omega f^p)^{1/p}$ .<sup>1</sup> The  $p \rightarrow q$  norm  $\|A\|_{p \rightarrow q}$  of a linear operator  $A$  between vector spaces of such functions is the smallest number  $c \geq 0$  such that  $\|Af\|_q \leq c\|f\|_p$  for all functions  $f$  in the domain of  $A$ . We also define the  $p \rightarrow q$  norm of a subspace  $V$  to be the maximum of  $\|f\|_q/\|f\|_p$  for  $f \in V$ ; note that for  $p = 2$  this is the same as the norm of the projector operator into  $V$ .

In this work, we are interested in the case  $p < q$  and we will call such  $p \rightarrow q$  norms *hypercontractive*.<sup>2</sup> Roughly speaking, for  $p < q$ , a function  $f$  with large  $\|f\|_q$  compared to  $\|f\|_p$  can be thought of as “spiky” or somewhat sparse (i.e., much of the mass concentrated in small portion of the entries). Hence finding a function  $f$  in a linear subspace  $V$  maximizing  $\|f\|_q/\|f\|_2$  for some  $q > 2$  can be thought of as a geometric analogue of the problem finding the shortest word in a linear code. This problem is equivalent to computing the  $2 \rightarrow q$  norm of the projector  $P$  into  $V$  (since  $\|Pf\|_2 \leq \|f\|_2$ ). Also when  $A$  is a normalized adjacency matrix of a graph (or more generally a Markov operator), upper bounds on the  $p \rightarrow q$  norm are known as *mixed-norm*, *Nash* or *hypercontractive inequalities* and can be used to show rapid mixing of the corresponding random walk (e.g., see the surveys [Gro75, SC97]). Such bounds also have many applications to theoretical computer science, which are described in the survey [Bis11].

However, very little is known about the complexity of computing these norms. This is in contrast to the case of  $p \rightarrow q$  norms for  $p \geq q$ , where much more is known both in terms of algorithms and lower bounds, see [Ste05, KNS08, BV11].

## 2 Our Results

We initiate a study of the computational complexity of approximating the  $2 \rightarrow 4$  (and more generally  $2 \rightarrow q$  for  $q > 2$ ) norm. While there are still many more questions than answers on this topic, we are able to show some new algorithmic and hardness results, as well as connections to both Khot’s unique games conjecture [Kho02] (UGC) and questions from quantum information theory. In particular our paper gives some conflicting evidence regarding the validity of the UGC and its close variant—the small set expansion hypothesis (SSEH) of [RS10]. (See also our conclusions section.)

First, we show in Theorem 2.5 that approximating the  $2 \rightarrow 4$  problem to within any constant factor cannot be done in polynomial time (unless SAT can be solved in  $\exp(o(n))$  time) but yet this problem is seemingly related to the UNIQUE GAMES and SMALL-SET EXPANSION problems. In particular, we show that approximating the  $2 \rightarrow 4$  norm is SMALL-SET EXPANSION-hard but yet has a subexponential algorithm which closely related to the [ABS10] algorithm for UNIQUE GAMES and SMALL-SET EXPANSION. Thus the computational difficulty of this problem can be considered as some indirect evidence *supporting* the validity of the UGC (or perhaps some weaker variants of it). To our knowledge, this is the first evidence of this kind for the UGC.

On the other hand, we show that a natural polynomial-time algorithm (based on an SDP hierarchy) that solves the previously proposed hard instances for UNIQUE GAMES.

---

<sup>1</sup> We follow the convention to use *expectation* norms for *functions* (on probability spaces) and *counting* norms, denoted as  $\|v\|_p = (\sum_{i=1}^n |v_i|^p)^{1/p}$ , for *vectors*  $v \in \mathbb{R}^n$ . All normed spaces here will be finite dimensional. We distinguish between expectation and counting norms to avoid recurrent normalization factors.

<sup>2</sup>We use this name because a bound of the form  $\|A\|_{p \rightarrow q} \leq 1$  for  $p < q$  is often called a *hypercontractive inequality*.

The previous best algorithms for some of these instances took almost exponential ( $\exp(\exp(\log^{\Omega(1)} n))$ ) time, and in fact they were shown to require super-polynomial time for some hierarchies. Thus this result suggests that this algorithm could potentially refute the UGC, and hence can be construed as evidence *opposing* the UGC’s validity.

## 2.1 Algorithms

We show several algorithmic results for the  $2 \rightarrow 4$  (and more generally  $2 \rightarrow q$ ) norm.

### 2.1.1 Subexponential algorithm for “good” approximation

For  $q \geq 2$ , we say that an algorithm provides a  $(c, C)$ -approximation for the  $2 \rightarrow q$  norm if on input an operator  $A$ , the algorithm can distinguish between the case that  $\|A\|_{2 \rightarrow q} \leq c\sigma$  and the case that  $\|A\|_{2 \rightarrow q} \geq C\sigma$ , where  $\sigma = \sigma_{\min}(A)$  is the minimum nonzero singular value of  $A$ . (Note that since we use the expectation norm,  $\|Af\|_q \geq \|Af\|_2 \geq \sigma\|f\|_2$  for every function  $f$  orthogonal to the Kernel of  $A$ .) We say that an algorithm provides a *good approximation* for the  $2 \rightarrow q$  norm if it provides a  $(c, C)$ -approximation for some (dimension independent) constants  $c < C$ . The motivation behind this definition is to capture the notion of a *dimension independent* approximation factor, and is also motivated by Theorem 2.4 below, that relates a good approximation for the  $2 \rightarrow q$  norm to solving the SMALL-SET EXPANSION problem.

We show the following:

**Theorem 2.1.** *For every  $1 < c < C$ , there is a  $\text{poly}(n) \exp(n^{2/q})$ -time algorithm that computes a  $(c, C)$ -approximation for the  $2 \rightarrow q$  norm of any linear operator whose range is  $\mathbb{R}^n$ .*

Combining this with our results below, we get as a corollary a subexponential algorithm for the SMALL-SET EXPANSION problem matching the parameters of [ABS10]’s algorithm. We note that this algorithm can be achieved by the “Sum of Squares” SDP hierarchy described below (and probably weaker hierarchies as well, although we did not verify this).

### 2.1.2 Polynomial algorithm for specific instances

We study a natural semidefinite programming (SDP) relaxation for computing the  $2 \rightarrow 4$  norm of a given linear operator which we call Tensor-SDP.<sup>3</sup> While Tensor-SDP is very unlikely to provide a poly-time constant-factor approximation for the  $2 \rightarrow 4$  norm in general (see Theorem 2.5 below), we do show that it provides such approximation on two very different types of instances:

- We show that Tensor-SDP certifies a constant upper bound on the ratio  $\|A\|_{2 \rightarrow 4} / \|A\|_{2 \rightarrow 2}$  where  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a *random linear operator* (e.g., obtained by a matrix with entries chosen as i.i.d Bernoulli variables) and  $m \geq \Omega(n^2 \log n)$ . In contrast, if  $m = o(n^2)$  then this ratio is  $\omega(1)$ , and hence this result is almost tight in the sense of obtaining “good approximation” in the sense mentioned above. We find this interesting, since random matrices seem like natural instances; indeed for superficially similar problems such shortest codeword, shortest lattice vector (or even the  $1 \rightarrow 2$  norm), it seems hard to efficiently certify bounds on random operators.

<sup>3</sup>We use the name Tensor-SDP for this program since it will be a canonical relaxation of the polynomial program  $\max_{\|x\|_2=1} \langle T, x^{\otimes 4} \rangle$  where  $T$  is the 4-tensor such that  $\langle T, x^{\otimes 4} \rangle = \|Ax\|_4^4$ . See Section 4.5 for more details.

- We show that Tensor-SDP gives a good approximation of the  $2 \rightarrow 4$  norm of the operator projecting a function  $f : \{\pm 1\}^n \rightarrow \mathbb{R}$  into its low-degree component:

**Theorem 2.2.** *Let  $\mathcal{P}_d$  be the linear operator that maps a function  $f : \{\pm 1\}^n \rightarrow \mathbb{R}$  of the form  $f = \sum_{\alpha \subseteq [n]} \hat{f}_\alpha \chi_\alpha$  to its low-degree part  $f' = \sum_{|\alpha| \leq d} \hat{f}_\alpha \chi_\alpha$  (where  $\chi_\alpha(x) = \prod_{i \in \alpha} x_i$ ). Then  $\text{Tensor-SDP}(\mathcal{P}_d) \leq 9^d$ .*

The fact that  $\mathcal{P}_d$  has bounded  $2 \rightarrow 4$  norm is widely used in the literature relating to the UGC. Previously, no general-purpose algorithm was known to efficiently certify this fact.

### 2.1.3 Quasipolynomial algorithm for additive approximation

We also consider the generalization of Tensor-SDP to a natural SDP *hierarchy*. This is a convex relaxation that starts from an initial SDP and tightens it by adding additional constraints. Such hierarchies are generally parameterized by a number  $r$  (often called the *number of rounds*), where the 1<sup>st</sup> round corresponds to the initial SDP, and the  $n^{\text{th}}$  round (for discrete problems where  $n$  is the instance size) corresponds to the exponential brute force algorithm that outputs an optimal answer. Generally, the  $r^{\text{th}}$ -round of each such hierarchy can be evaluated in  $n^{O(r)}$  time (though in some cases  $n^{O(1)}2^{O(r)}$  time suffices [BRS11]). See Section 3, as well as the surveys [CT10, Lau03] and the papers [SA90, LS91, RS09, KPS10] for more information about these hierarchies.

We call the hierarchy we consider here the *Sum of Squares* (SoS) hierarchy. It is not novel but rather a variant of the hierarchies studied by several authors including Shor [Sho87], Parrilo [Par00, Par03], Nesterov [Nes00] and Lasserre [Las01]. (Generally in our context these hierarchies can be made equivalent in power, though there are some subtleties involved; see [Lau09] and Appendix C for more details.) We describe the SoS hierarchy formally in Section 3. We show that Tensor-SDP’s extension to several rounds of the SoS hierarchy gives a non-trivial *additive* approximation:

**Theorem 2.3.** *Let  $\text{Tensor-SDP}^{(d)}$  denote the  $n^{O(d)}$ -time algorithm by extending Tensor-SDP to  $d$  rounds of the Sum-of-Squares hierarchy. Then for all  $\varepsilon$ , there is  $d = O(\log(n)/\varepsilon^2)$  such that*

$$\|A\|_{2 \rightarrow 4}^4 \leq \text{Tensor-SDP}^{(d)}(A) \leq \|A\|_{2 \rightarrow 4}^4 + \varepsilon \|A\|_{2 \rightarrow 2}^2 \|A\|_{2 \rightarrow \infty}^2.$$

The term  $\|A\|_{2 \rightarrow 2}^2 \|A\|_{2 \rightarrow \infty}^2$  is a natural upper bound on  $\|A\|_{2 \rightarrow 4}^4$  obtained using Hölder’s inequality. Since  $\|A\|_{2 \rightarrow 2}$  is the largest singular value of  $A$ , and  $\|A\|_{2 \rightarrow \infty}$  is the largest 2-norm of any row of  $A$ , they can be computed quickly. Theorem 2.3 shows that one can improve this upper bound by a factor of  $\varepsilon$  using run time  $\exp(\log^2(n)/\varepsilon^2)$ . Note however that in the special case (relevant to the UGC) that  $A$  is a projector to a subspace  $V$ ,  $\|A\|_{2 \rightarrow 2} = 1$  and  $\|A\|_{2 \rightarrow \infty} \geq \sqrt{\dim(V)}$  (see Lemma 10.1), which unfortunately means that Theorem 2.3 does not give any new algorithms in that setting.

Despite Theorem 2.3 being a non-quantum algorithm for for an ostensibly non-quantum problem, we actually achieve it using the results of Brandão, Christiandl and Yard [BaCY11] about the quantum separability problem. In fact, it turns out that the SoS hierarchy extension of Tensor-SDP is equivalent to techniques that have been used to approximate separable states [DPS04]. We find this interesting both because there are few positive general results about the convergence rate of SDP hierarchies, and because the techniques of [BaCY11], based on entanglement measures of quantum states, are different from typical ways of proving correctness of semidefinite programs, and in particular different techniques

from the ones we use to analyze Tensor-SDP in other settings. This connection also means that integrality gaps for Tensor-SDP would imply new types of separable states that pass most of the known tests for entanglement.

## 2.2 Reductions

We relate the question of computing the hypercontractive norm with two other problems considered in the literature: the *small set expansion* problem [RS10, RST10a], and the *injective tensor norm* question studied in the context of quantum information theory [HM10, BaCY11].

### 2.2.1 Hypercontractivity and small set expansion

Khot’s *Unique Games Conjecture* [Kho02] (UGC) has been the focus of intense research effort in the last few years. The conjecture posits the hardness of approximation for a certain constraint-satisfaction problem, and shows promise to settle many open questions in the theory of approximation algorithms. Many works have been devoted to studying the plausibility of the UGC, as well as exploring its implications and obtaining unconditional results inspired or motivated by this effort. Tantalizingly, at the moment we have very little insight on whether this conjecture is actually true, and thus producing evidence on the UGC’s truth or falsity is a central effort in computational complexity. Raghavendra and Steurer [RS10] proposed a hypothesis closely related to the UGC called the *Small-Set Expansion* hypothesis (SSEH). Loosely speaking, the SSEH states that it is NP-hard to certify that a given graph  $G = (V, E)$  is a *small-set expander* in the sense that subsets with size  $o(|V|)$  vertices have almost all their neighbors outside the set. [RS10] showed that SSEH implies UGC. While a reduction in the other direction is not known, all currently known algorithmic and integrality gap results apply to both problems equally well (e.g., [ABS10, RST10b]), and thus the two conjectures are likely to be equivalent.

We show, loosely speaking, that a graph is a small-set expander if and only if the projection operator to the span of its top eigenvectors has bounded  $2 \rightarrow 4$  norm. To make this precise, if  $G = (V, E)$  is a regular graph, then let  $P_{\geq \lambda}(G)$  be the projection operator into the span of the eigenvectors of  $G$ ’s normalized adjacency matrix with eigenvalue at least  $\lambda$ , and  $\Phi_G(\delta)$  be  $\min_{S \subseteq V, |S| \leq \delta|V|} \mathbb{P}_{(u,v) \in E} [v \notin S | u \in S]$ .

Then we relate small-set expansion to the  $2 \rightarrow 4$  norm (indeed the  $2 \rightarrow q$  norm for even  $q \geq 4$ ) as follows:

**Theorem 2.4.** *For every regular graph  $G$ ,  $\lambda > 0$  and even  $q$ ,*

1. (Norm bound implies expansion) *For all  $\delta > 0, \varepsilon > 0$ ,  $\|P_{\geq \lambda}(G)\|_{2 \rightarrow q} \leq \varepsilon / \delta^{(q-2)/2q}$  implies that  $\Phi_G(\delta) \geq 1 - \lambda - \varepsilon^2$ .*
2. (Expansion implies norm bound) *There is a constant  $c$  such that for all  $\delta > 0$ ,  $\Phi_G(\delta) > 1 - \lambda 2^{-cq}$  implies  $\|P_{\geq \lambda}(G)\|_{2 \rightarrow q} \leq 2 / \sqrt{\delta}$ .*

While one direction (bounded hypercontractive norm implies small-set expansion) was already known,<sup>4</sup> to our knowledge the other direction is novel. As a corollary we show that the SSEH implies that there is no good approximation for the  $2 \rightarrow 4$  norm.

---

<sup>4</sup>While we do not know who was the first to point out this fact explicitly, within theoretical CS it was implicitly used in several results relating the Bonami-Beckner-Gross hypercontractivity of the Boolean noise operator to isoperimetric properties, with one example being O’Donnell’s proof of the soundness of [KV05]’s integrality gap (see [KV05, Sec 9.1]).



### 2.2.2 Hypercontractivity and the injective tensor norm

We are able to make progress in understanding both the complexity of the  $2 \rightarrow 4$  norm and the quality of our SDP relaxation by relating the  $2 \rightarrow 4$  norm to several natural questions about tensors. An  $r$ -tensor can be thought of as an  $r$ -linear form on  $\mathbb{R}^n$ , and the *injective tensor norm*  $\|\cdot\|_{\text{inj}}$  of a tensor is given by maximizing this form over all unit vector inputs. See Section 9 for a precise definition. When  $r = 1$ , this norm is the 2-norm of a vector and when  $r = 2$ , it is the operator norm (or  $2 \rightarrow 2$ -norm) of a matrix, but for  $r = 3$  it becomes NP-hard to calculate. One motivation to study this norm comes from quantum mechanics, where computing it is equivalent to a number of long-studied problems concerning entanglement and many-body physics [HM10]. More generally, tensors arise in a vast range of practical problems involving multidimensional data [vL09] for which the injective tensor norm is both of direct interest and can be used as a subroutine for other tasks, such as tensor decomposition [dlVKKV05].

It is not hard to show that  $\|A\|_{2 \rightarrow 4}^4$  is actually equal to  $\|T\|_{\text{inj}}$  for some 4-tensor  $T = T_A$ . Not all 4-tensors can arise this way, but we show that the injective tensor norm problem for general tensors can be reduced to those of the form  $T_A$ . Combined with known results about the hardness of tensor computations, this reduction implies the following hardness result. To formulate the theorem, recall that the Exponential Time Hypothesis (ETH) [IPZ98] states that 3-SAT instances of length  $n$  require time  $\exp(\Omega(n))$  to solve.

**Theorem 2.5** (informal version). *Assuming ETH, then for any  $\varepsilon, \delta$  satisfying  $2\varepsilon + \delta < 1$ , the  $2 \rightarrow 4$  norm of an  $m \times m$  matrix  $A$  cannot be approximated to within a  $\exp(\log^\varepsilon(m))$  multiplicative factor in time less than  $m^{\log^\delta(m)}$  time. This hardness result holds even with  $A$  is a projector.*

While we are primarily concerned with the case of  $\Omega(1)$  approximation factor, we note that poly-time approximations to within multiplicative factor  $1 + 1/n^{1.01}$  are not possible unless  $\mathbf{P} = \mathbf{NP}$ . This, along with Theorem 2.5, is restated more formally as Theorem 9.4 in Section 9.2. We also note that Theorem 2.5 yields as a corollary that, assuming ETH, there is no polynomial-time algorithm obtaining a good approximation for the  $2 \rightarrow 4$  norm. We note that these results hold under weaker assumptions than the ETH; see Section 9.2 as well.

Previously no hardness results were known for the  $2 \rightarrow 4$  norm, or any  $p \rightarrow q$  norm with  $p < q$ , even for calculating the norms exactly. However, hardness of approximation results for  $1 + 1/\text{poly}(n)$  multiplicative error have been proved for other polynomial optimization problems [BTN98].

### 2.3 Relation to the Unique Games Conjecture

Our results and techniques have some relevance to the unique games conjecture. Theorem 2.4 shows that obtaining a good approximation for the  $2 \rightarrow q$  norm is SMALL-SET EXPANSION hard, but Theorem 2.1 shows that this problem is not “that much harder” than UNIQUE GAMES and SMALL-SET EXPANSION since it too has a subexponential algorithm. Thus, the  $2 \rightarrow q$  problem is in some informal sense “of similar flavor” to the UNIQUE GAMES/SMALL-SET EXPANSION. On the other hand, we actually are able to show in Theorem 2.5 *hardness* (even if only quasipolynomial) to this problem, whereas a similar result for UNIQUE GAMES or SMALL-SET EXPANSION would be a major breakthrough. So there is a sense in which these results can be thought of as some “positive evidence” in favor of at least weak variants of the UGC. (We emphasize however that there are inherent difficulties in extending these

results for UNIQUE GAMES, and it may very well be that obtaining a multiplicative approximation to the  $2 \rightarrow 4$  of an operator is significantly harder problem than UNIQUE GAMES or SMALL-SET EXPANSION.) In contrast, our positive algorithmic results show that perhaps the  $2 \rightarrow q$  norm can be thought of as a path to refuting the UGC. In particular we are able to extend our techniques to show a polynomial time algorithm can approximate the canonical hard instances for UNIQUE GAMES considered in prior works.

**Theorem 2.6.** *(Informal) Eight rounds of the SoS relaxation certifies that it is possible to satisfy at most 1/100 fraction of the constraints of UNIQUE GAMES instances of the “quotient noisy cube” and “short code” types considered in [RS09, KS09, KPS10, BGH<sup>+</sup>11]*

These instances are the same ones for which previous works showed that weaker hierarchies such as “SDP+Sherali Adams” and “Approximate Lasserre” require  $\omega(1)$  rounds to certify that one cannot satisfy almost all the constraints [KV05, RS10, KS09, BGH<sup>+</sup>11]. In fact, for the “short code” based instances of [BGH<sup>+</sup>11] there was no upper bound known better than  $\exp(\log^{\Omega(1)} n)$  on the number of rounds required to certify that they are not almost satisfiable, regardless of the power of the hierarchy used.

This is significant since the current best known algorithms for UNIQUE GAMES utilize SDP hierarchies [BRS11, GS11],<sup>5</sup> and the instances above were the only known evidence that polynomial time versions of these algorithms do not refute the unique games conjecture. Our work also show that strong “basis independent” hierarchies such as Sum of Squares [Par00, Par03] and Lasserre [Las01] can in fact do better than the seemingly only slightly weaker variants.<sup>6</sup>

### 3 The SoS hierarchy

For our algorithmic results in this paper we consider a semidefinite programming (SDP) hierarchy that we call the *Sum of Squares* (SoS) hierarchy. We call the hierarchy we consider here the *Sum of Squares* (SoS) hierarchy. This is not a novel algorithm and essentially the same hierarchies were considered by many other researchers (see the survey [Lau09]). Because different works sometimes used slightly different definitions, in this section we formally define the hierarchy we use as well as explain the intuition behind it. While there are some subtleties involved, one can think of this hierarchy as equivalent in power to the programs considered by Parrilo, Lasserre and others, while stronger than hierarchies such “SDP+Sherali-Adams” and “Approximate Lasserre” considered in [RS09, KPS10, BRS11].

The SoS SDP is a relaxation for polynomial equations. That is, we consider a system of the following form: maximize  $P_0(x)$  over  $x \in \mathbb{R}^n$  subject to  $P_i^2(x) = 0$  for  $i = 1 \dots m$  and  $P_0, \dots, P_m$  polynomials of degree at most  $d$ .<sup>7</sup> For  $r \geq 2d$ , the  $r$ -round SoS SDP optimizes over  $x_1, \dots, x_n$  that can be thought of as formal variables rather than actual numbers. For these formal variables, expressions of the form  $P(x)$  are well defined and correspond to a

<sup>5</sup>Both these works showed SDP-hierarchy-based algorithms matching the performance of the subexponential algorithm of [ABS10]. [GS11] used the Lasserre hierarchy, while [BRS11] used the weaker “SDP+Sherali-Adams” hierarchy.

<sup>6</sup>The only other result of this kind we are aware of is [KMN11], that show that Lasserre gives a better approximation ratio than the linear programming Sherali-Adams hierarchy for the knapsack problem. We do not know if weaker semidefinite hierarchies match this ratio, although knapsack of course has a simple dynamic programming based PTAS.

<sup>7</sup>This form is without loss of generality, as one can translate an inequality constraint of the form  $P_i(x) \geq 0$  into the equality constraint  $(P_i(x) - y^2)^2 = 0$  where  $y$  is some new auxiliary variable. It is useful to show equivalences between various hierarchy formulations; see also Appendix C.

real number (which can be computed from the SDP solution) as long as  $P$  is a polynomial of degree at most  $r$ . These numbers obey the *linearity* property which is that  $(P+Q)(x) = P(x)+Q(x)$ , and, most importantly, the *positivity* property that  $P^2(x) \geq 0$  for every polynomial  $P$  of degree at most  $r/2$ . These expressions satisfy all initial constraints (i.e.,  $P_i^2(x) = 0$  for  $i = 1 \dots m$ ) and the *value* of the SDP is set to be the expression  $P_0(x)$ . The above means that to show that the SoS relaxation has value at most  $v$  it is sufficient to give any proof that derives from the constraints  $\{P_i^2(x) = 0\}_{i=1..m}$  the conclusion that  $P_0(x) \leq v$  using only the linearity and positivity properties, without using any polynomials of degree larger than  $r$  in the intermediate steps. In fact, such a proof always has the form

$$v - P_0(x) = \sum_{i=1}^k R_i(x)^2 + \sum_{i=1}^m P_i(x)Q_i(x), \quad (3.1)$$

where  $R_1, \dots, R_k, Q_1, \dots, Q_m$  are arbitrary polynomials satisfying  $\deg R_i \leq r/2, \deg P_i Q_i \leq r$ . The polynomial  $\sum_i R_i(x)^2$  is a SoS (sum of squares) and optimizing over such polynomials (along with the  $Q_1, \dots, Q_m$ ) can be achieved with a semi-definite program.

**Pseudo-expectation view.** For more intuition about the SoS hierarchy, one can imagine that instead of being formal variables,  $x_1, \dots, x_n$  actually correspond to correlated random variables  $X_1, \dots, X_n$  over  $\mathbb{R}^n$ , and the expression  $P(x)$  is set to equal the expectation  $\mathbb{E}[P(X)]$ . In this case, the linearity and positivity properties are obviously satisfied by these expressions, although other properties that would be obtained if  $x_1, \dots, x_n$  were simply numbers might not hold. For example, the property that  $R(x) = P(x)Q(x)$  if  $R = P \cdot Q$  does not necessarily hold, since its not always the case that  $E[XY] = E[X]E[Y]$  for every three random variables  $X, Y, Z$ . So, another way to describe the  $r$ -round SoS hierarchy is that the expressions  $P(x)$  (for  $P$  of degree at most  $r$ ) satisfy some of the constraints that would have been satisfied if these expressions corresponded to expectations over some correlated random variables  $X_1, \dots, X_N$ . For this reason, we will use the notation  $\tilde{\mathbb{E}}_x P(x)$  instead of  $P(x)$  where we refer to the functional  $\tilde{\mathbb{E}}$  as a level- $r$  *pseudo-expectation functional* (or  $r$ -p.e.f. for short). Also, rather than describing  $x_1, \dots, x_n$  as formal variables, we will refer to them as level- $r$  *fictitious random variables* (or  $r$ -f.r.v. for short) since in some sense they look like true correlated random variables up to their  $r^{\text{th}}$  moment.

We can now present our formal definition of pseudo-expectation and the SoS hierarchy:<sup>8</sup>

**Definition 3.1.** Let  $\tilde{\mathbb{E}}$  be a functional that maps polynomial  $P$  over  $\mathbb{R}^n$  of degree at most  $r$  into a real number which we denote by  $\tilde{\mathbb{E}}_x P(x)$  or  $\tilde{\mathbb{E}} P$  for short. We say that  $\tilde{\mathbb{E}}$  is a *level- $r$  pseudo-expectation functional* ( $r$ -p.e.f. for short) if it satisfies:

**Linearity** For every polynomials  $P, Q$  of degree at most  $r$  and  $\alpha, \beta \in \mathbb{R}$ ,  $\tilde{\mathbb{E}}(\alpha P + \beta Q) = \alpha \tilde{\mathbb{E}} P + \beta \tilde{\mathbb{E}} Q$ .

**Positivity** For every polynomial  $P$  of degree at most  $r/2$ ,  $\tilde{\mathbb{E}} P^2 \geq 0$ .

**Normalization**  $\tilde{\mathbb{E}} 1 = 1$  where on the RHS, 1 denotes the degree-0 polynomial that is the constant 1.

**Definition 3.2.** Let  $P_0, \dots, P_m$  be polynomials over  $\mathbb{R}^n$  of degree at most  $d$ , and let  $r \geq 2d$ . The value of the  $r$ -round SoS SDP for the program “max  $P_0$  subject to  $P_i^2 = 0$  for

<sup>8</sup>We use the name “Sum of Squares” since the positivity condition below is the most important constraint of this program. However, some prior works used this name for the *dual* of the program we define here. As we show in Appendix C, in many cases of interest to us there is no duality gap.

$i = 1 \dots m$ ", is equal to the maximum of  $\tilde{\mathbb{E}} P_0$  where  $\tilde{\mathbb{E}}$  ranges over all level  $r$  pseudo-expectation functionals satisfying  $\tilde{\mathbb{E}} P_i^2 = 0$  for  $i = 1 \dots m$ .

The functional  $\tilde{\mathbb{E}}$  can be represented by a table of size  $n^{O(r)}$  containing the pseudo-expectations of every monomial of degree at most  $r$  (or some other linear basis for polynomials of degree at most  $r$ ). For a linear functional  $\tilde{\mathbb{E}}$ , the map  $P \mapsto \tilde{\mathbb{E}} P^2$  is a quadratic form. Hence,  $\tilde{\mathbb{E}}$  satisfies the positivity condition if and only if the corresponding quadratic form is positive semidefinite. It follows that the convex set of level- $r$  pseudo-expectation functionals over  $\mathbb{R}^n$  admits an  $n^{O(r)}$ -time separation oracle, and hence the  $r$ -round SoS relaxation can be solved up to accuracy  $\varepsilon$  in time  $(mn \cdot \log(1/\varepsilon))^{O(r)}$ .

As noted above, for every random variable  $X$  over  $\mathbb{R}^n$ , the functional  $\tilde{\mathbb{E}} P := \mathbb{E} P(X)$  is a level- $r$  pseudo-expectation functional for every  $r$ . As  $r \rightarrow \infty$ , this hierarchy of pseudo-expectations will converge to the expectations of a true random variable [Las01], but the convergence is in general not guaranteed to happen in a finite number of steps [dKL11].

Whenever there can be ambiguity about what are the variables of the polynomial  $P$  inside an  $r$ -p.e.f.  $\tilde{\mathbb{E}}$ , we will use the notation  $\tilde{\mathbb{E}}_x P(x)$  (e.g.,  $\tilde{\mathbb{E}}_x x_3^2$  is the same as  $\tilde{\mathbb{E}} P$  where  $P$  is the polynomial  $x \mapsto x_3^2$ ). As mentioned above, we call the inputs  $x$  to the polynomial *level- $r$  fictitious random variables* or  *$r$ -f.r.v.* for short.

**Remark 3.3.** The main difference between the SoS hierarchy and weaker SDP hierarchies considered in the literature such as SDP+Sherali Adams and the Approximate Lasserre hierarchies [RS09, KPS10] is that the SoS hierarchy treats all polynomials equally and hence is agnostic to the choice of basis. For example, the approximate Lasserre hierarchy can also be described in terms of pseudo-expectations, but these pseudo-expectations are only defined for monomials, and are allowed some small error. While they can be extended linearly to other polynomials, for non-sparse polynomials that error can greatly accumulate.

### 3.1 Basic properties of pseudo-expectation

For two polynomials  $P$  and  $Q$ , we write  $P \leq Q$  if  $Q = P + \sum_{i=1}^m R_i^2$  for some polynomials  $R_1, \dots, R_m$ .

If  $P$  and  $Q$  have degree at most  $r$ , then  $P \leq Q$  implies that  $\tilde{\mathbb{E}} P \leq \tilde{\mathbb{E}} Q$  every  $r$ -p.e.f.  $\tilde{\mathbb{E}}$ . This follows using linearity and positivity, as well as the (not too hard to verify) observation that if  $Q - P = \sum_i R_i^2$  then it must hold that  $\deg(R_i) \leq \max\{\deg(P), \deg(Q)\}/2$  for every  $i$ .

We would like to understand how polynomials behave on linear subspaces of  $\mathbb{R}^n$ . A map  $P: \mathbb{R}^n \rightarrow \mathbb{R}$  is *polynomial* over a linear subspace  $V \subseteq \mathbb{R}^n$  if  $P$  restricted to  $V$  agrees with a polynomial in the coefficients for some basis of  $V$ . Concretely, if  $g_1, \dots, g_m$  is an (orthonormal) basis of  $V$ , then  $P$  is *polynomial* over  $V$  if  $P(f)$  agrees with a polynomial in  $\langle f, g_1 \rangle, \dots, \langle f, g_m \rangle$ . We say that  $P \leq Q$  holds over a subspace  $V$  if  $P - Q$ , as a polynomial over  $V$ , is a sum of squares.

**Lemma 3.4.** *Let  $P$  and  $Q$  be two polynomials over  $\mathbb{R}^n$  of degree at most  $r$ , and let  $B: \mathbb{R}^n \rightarrow \mathbb{R}^k$  be a linear operator. Suppose that  $P \leq Q$  holds over the kernel of  $B$ . Then,  $\tilde{\mathbb{E}} P \leq \tilde{\mathbb{E}} Q$  holds for any  $r$ -p.e.f.  $\tilde{\mathbb{E}}$  over  $\mathbb{R}^n$  that satisfies  $\tilde{\mathbb{E}}_f \|Bf\|^2 = 0$ .*

*Proof.* Since  $P \leq Q$  over the kernel of  $B$ , we can write  $Q(f) = P(f) + \sum_{i=1}^m R_i^2(f) + \sum_{j=1}^k (Bf)_j S_j(f)$  for polynomials  $R_1, \dots, R_m$  and  $S_1, \dots, S_k$  over  $\mathbb{R}^n$ . By positivity,  $\tilde{\mathbb{E}}_f R_i^2(f) \geq 0$  for all  $i \in [m]$ . We claim that  $\tilde{\mathbb{E}}_f (Bf)_j S_j(f) = 0$  for all  $j \in [k]$  (which would finish the proof). This claim follows from the fact that  $\tilde{\mathbb{E}}_f (Bf)_j^2 = 0$  for all  $j \in [k]$  and Lemma 3.5 below.  $\square$

**Lemma 3.5** (Pseudo Cauchy-Schwarz). *Let  $P$  and  $Q$  be two polynomials of degree at most  $r$ . Then,  $\tilde{\mathbb{E}} PQ \leq \sqrt{\tilde{\mathbb{E}} P^2} \cdot \sqrt{\tilde{\mathbb{E}} Q^2}$  for any degree- $2r$  pseudo-expectation functional  $\tilde{\mathbb{E}}$ .*

*Proof.* We first consider the case  $\tilde{\mathbb{E}} P^2, \tilde{\mathbb{E}} Q^2 > 0$ . Then, by linearity of  $\tilde{\mathbb{E}}$ , we may assume that  $\tilde{\mathbb{E}} P^2 = \tilde{\mathbb{E}} Q^2 = 1$ . Since  $2PQ \leq P^2 + Q^2$  (by expanding the square  $(P - Q)^2$ ), it follows that  $\tilde{\mathbb{E}} PQ \leq \frac{1}{2} \tilde{\mathbb{E}} P^2 + \frac{1}{2} \tilde{\mathbb{E}} Q^2 = 1$  as desired. It remains to consider the case  $\tilde{\mathbb{E}} P^2 = 0$ . In this case,  $2\alpha PQ \leq P^2 + \alpha^2 Q^2$  implies that  $\tilde{\mathbb{E}} PQ \leq \alpha \cdot \frac{1}{2} \tilde{\mathbb{E}} Q^2$  for all  $\alpha > 0$ . Thus  $\tilde{\mathbb{E}} PQ = 0$ , as desired.  $\square$

Lemma 3.5 also explains why our SDP in Definition 3.2 is dual to the one in (3.1). If  $\tilde{\mathbb{E}}$  is a level- $r$  pseudo-expectation functional satisfying  $\tilde{\mathbb{E}}[P_i^2] = 0$ , then Lemma 3.5 implies that  $\tilde{\mathbb{E}}[P_i Q_i] = 0$  for all  $Q_i$  with  $\deg P_i Q_i \leq r$ .

Appendix A contains some additional useful facts about pseudo-expectation functionals. In particular, we will make repeated use of the fact that they satisfy another Cauchy-Schwarz analogue: namely, for any level-2 f.r.v.'s  $f, g$ , we have  $\tilde{\mathbb{E}}_{f,g} \langle f, g \rangle \leq \sqrt{\tilde{\mathbb{E}}_f \|f\|^2} \sqrt{\tilde{\mathbb{E}}_g \|g\|^2}$ . This is proven in Lemma A.4.

### 3.2 Why is this SoS hierarchy useful?

Consider the following example. It is known that if  $f : \{\pm 1\}^\ell \rightarrow \mathbb{R}$  is a degree- $d$  polynomial then

$$9^d \left( \mathbb{E}_{w \in \{\pm 1\}^\ell} f(w)^2 \right)^2 \geq \mathbb{E}_{w \in \{\pm 1\}^\ell} f(w)^4, \quad (3.2)$$

(see e.g. [O'D07]). Equivalently, the linear operator  $\mathcal{P}_d$  on  $\mathbb{R}^{\{\pm 1\}^\ell}$  that projects a function into the degree  $d$  polynomials satisfies  $\|\mathcal{P}_d\|_{2 \rightarrow 4} \leq 9^{d/4}$ . This fact is known as the hypercontractivity of low-degree polynomials, and was used in several integrality gaps results such as [KV05]. By following the proof of (3.2) we show in Lemma 5.1 that a stronger statement is true:

$$9^d \left( \mathbb{E}_{w \in \{\pm 1\}^\ell} f(w)^2 \right)^2 = \mathbb{E}_{w \in \{\pm 1\}^\ell} f(w)^4 + \sum_{i=1}^m Q_i(f)^2, \quad (3.3)$$

where the  $Q_i$ 's are polynomials of degree  $\leq 2$  in the  $\binom{\ell}{d}$  variables  $\{\hat{f}(\alpha)\}_{\alpha \in \binom{[\ell]}{d}}$  specifying the coefficients of the polynomial  $f$ . By using the positivity constraints, (3.3) implies that (3.2) holds even in the 4-round SoS relaxation where we consider the coefficients of  $f$  to be given by 4-f.r.v. This proves Theorem 2.2, showing that the SoS relaxation certifies that  $\|\mathcal{P}_d\|_{2 \rightarrow 4} \leq 9^{d/4}$ .

**Remark 3.6.** Unfortunately to describe the result above, we needed to use the term ‘‘degree’’ in two different contexts. The SDP relaxation considers polynomial expressions of degree at most 4 *in the coefficients of  $f$* . This is a different notion of degree than the degree  $d$  of  $f$  itself as a polynomial over  $\mathbb{R}^\ell$ . In particular the variables of this SoS program are the  $\binom{\ell}{d}$  coefficients  $\{\hat{f}(\alpha)\}_{\alpha \in \binom{[\ell]}{d}}$ . Note that for every fixed  $w$ , the expression  $f(w)$  is a linear polynomial over these variables, and hence the expressions  $\left( \mathbb{E}_{w \in \{\pm 1\}^\ell} f(w)^2 \right)^2$  and  $\mathbb{E}_{w \in \{\pm 1\}^\ell} f(w)^4$  are degree 4 polynomials over the variables.

While the proof of (3.3) is fairly simple, we find the result— that hypercontractivity of polynomials is efficiently certifiable— somewhat surprising. The reason is that hypercontractivity serves as the basis of the integrality gaps results which are exactly instances

of maximization problems where the objective value is low but this is supposedly hard to certify. In particular, we consider integrality gaps for `UNIQUE GAMES` considered before in the literature. All of these instances follow the framework initiated by Khot and Vishnoi [KV05]. Their idea was inspired by `UNIQUE GAMES` hardness proofs, with the integrality gap obtained by composing an initial instance with a gadget. The proof that these instances have “cheating” SDP solutions is obtained by “lifting” the completeness proof of the gadget. On the other hand, the soundness property of the gadget, combined with some isoperimetric results, showed that the instances do not have real solutions. This approach of lifting completeness proofs of reductions was used to get other integrality gap results as well [Tul09]. We show that the SoS hierarchy allows us to lift a certain *soundness* proof for these instances, which includes a (variant of) the invariance principle of [MOO05], influence-decoding ala [KKMO04], and hypercontractivity of low-degree polynomials. It turns out all these results can be proven via sum-of-squares type arguments and hence lifted to the SoS hierarchy.

## 4 Overview of proofs

We now give a very high level overview of the tools we use to obtain our results, leaving details to the later sections and appendices.

### 4.1 Subexponential algorithm for the 2-to-q norm

Our subexponential algorithm for obtaining a good approximation for the  $2 \rightarrow q$  norm is extremely simple. It is based on the observation that a subspace  $V \subseteq \mathbb{R}^n$  of too large a dimension must contain a function  $f$  such that  $\|f\|_q \gg \|f\|_2$ . For example, if  $\dim(V) \gg \sqrt{n}$ , then there must be  $f$  such that  $\|f\|_4 \gg \|f\|_2$ . This means that if we want to distinguish between, say, the case that  $\|V\|_{2 \rightarrow 4} \leq 2$  and  $\|V\|_{2 \rightarrow 4} \geq 3$ , then we can assume without loss of generality that  $\dim(V) = O(\sqrt{n})$  in which case we can solve the problem in  $\exp(O(\sqrt{n}))$  time. To get intuition, consider the case that  $V$  is spanned by an orthonormal basis  $f^1, \dots, f^d$  of functions whose entries are all in  $\pm 1$ . Then clearly we can find coefficients  $a_1, \dots, a_d \in \{\pm 1\}$  such that the first coordinate of  $g = \sum a_j f^j$  is equal to  $d$ , which means that its 4-norm is at least  $(d^4/n)^{1/4} = d/n^{1/4}$ . On the other hand, since the basis is orthonormal, the 2-norm of  $g$  equals  $\sqrt{d}$  which is  $\ll d/n^{1/4}$  for  $d \gg \sqrt{n}$ .

Note the similarity between this algorithm and [ABS10]’s algorithm for `SMALL-SET EXPANSION`, that also worked by showing that if the dimension of the top eigenspace of a graph is too large then it cannot be a small-set expander. Indeed, using our reduction of `SMALL-SET EXPANSION` to the  $2 \rightarrow q$  norm, we can reproduce a similar result to [ABS10].

### 4.2 Bounding the value of SoS relaxations

We show that in several cases, the SoS SDP hierarchy gives strong bounds on various instances. At the heart of these results is a general approach of “lifting” proofs about one-dimensional objects into the SoS relaxation domain. Thus we transform the prior proofs that these instances have small objective value, into a proof that the SoS relaxation also has a small objective. The crucial observation is that many proofs boil down to the simple fact that a sum of squares of numbers is always non-negative. It turns out that this “sum of squares” axiom is surprisingly powerful (e.g. implying a version of the Cauchy–Schwarz inequality given by Lemma A.4), and many proofs boil down to essentially this principle.

### 4.3 The 2-to-4 norm and small-set expansion

Bounds on the  $p \rightarrow q$  norm of operators for  $p < q$  have been used to show fast convergence of Markov chains. In particular, it is known that if the projector to the top eigenspace of a graph  $G$  has bounded  $2 \rightarrow 4$  norm, then that graph is a *small-set expander* in the sense that sets of  $o(1)$  fraction of the vertices have most of their edges exit the set. In this work we show a converse to this statement, proving that if  $G$  is a small-set expander, then the corresponding projector has bounded  $2 \rightarrow 4$  norm. As mentioned above, one corollary of this result is that a good (i.e., dimension-independent) approximation to the  $2 \rightarrow 4$  norm will refute the Small-Set Expansion hypothesis of [RS10].

We give a rough sketch of the proof. Suppose that  $G$  is a sufficiently strong small-set expander, in the sense that every set  $S$  with  $|S| \leq \delta|V(G)|$  has all but a tiny fraction of the edges  $(u, v)$  with  $u \in S$  satisfying  $v \notin S$ . Let  $f$  be a function in the eigenspace of  $G$  corresponding to eigenvalues larger than, say 0.99. Since  $f$  is in the top eigenspace, for the purposes of this sketch let's imagine that it satisfies

$$\forall x \in V, \mathbb{E}_{y \sim x} f(y) \geq 0.9f(x), \quad (4.1)$$

where the expectation is over a random neighbor  $y$  of  $x$ . Now, suppose that  $\mathbb{E} f(x)^2 = 1$  but  $\mathbb{E} f(x)^4 = C$  for some  $C \gg \text{poly}(1/\delta)$ . That means that most of the contribution to the 4-norm comes from the set  $S$  of vertices  $x$  such that  $f(x) \geq (1/2)C^{1/4}$ , but  $|S| \ll \delta|V(G)|$ . Moreover, suppose for simplicity that  $f(x) \in ((1/2)C^{1/4}, 2C^{1/4})$ , in which case the condition (\*) together with the small-set expansion condition that for most vertices  $y$  in  $\Gamma(S)$  (the neighborhood of  $S$ ) satisfy  $f(y) \geq C^{1/4}/3$ , but the small-set expansion condition, together with the regularity of the graph imply that  $|\Gamma(S)| \geq 200|S|$  (say), which implies that  $\mathbb{E} f(x)^4 \geq 2C$ —a contradiction.

The actual proof is more complicated, since we can't assume the condition (4.1). Instead we will approximate it by assuming that  $f$  is the function in the top eigenspace that *maximizes* the ratio  $\|f\|_4/\|f\|_2$ . See Section 8 for the details.

### 4.4 The 2-to-4 norm and the injective tensor norm

To relate the  $2 \rightarrow 4$  norm to the injective tensor norm, we start by establishing equivalences between the  $2 \rightarrow 4$  norm and a variety of different tensor problems. Some of these are straightforward exercises in linear algebra, analogous to proving that the largest eigenvalue of  $M^T M$  equals the square of the operator norm of  $M$ .

One technically challenging reduction is between the problem of optimizing a general degree-4 polynomial  $f(x)$  for  $x \in \mathbb{R}^n$  and a polynomial that can be written as the sum of fourth powers of linear functions of  $x$ . Straightforward approaches will magnify errors by  $\text{poly}(n)$  factors, which would make it impossible to rule out a PTAS for the  $2 \rightarrow 4$  norm. This would still be enough to prove that a  $1/\text{poly}(n)$  additive approximation is **NP**-hard. However, to handle constant-factor approximations, we will instead use a variant of a reduction in [HM10]. This will allow us to map a general tensor optimization problem (corresponding to a general degree-4 polynomial) to a  $2 \rightarrow 4$  norm calculation without losing very much precision.

To understand this reduction, we first introduce the  $n^2 \times n^2$  matrix  $A_{2,2}$  (defined in Section 9) with the property that  $\|A\|_{2 \rightarrow 4}^4 = \max z^T A_{2,2} z$ , where the maximum is taken over unit vectors  $z$  that can be written in the form  $x \otimes y$ . Without this last restriction, the maximum would simply be the operator norm of  $A_{2,2}$ . Operationally, we can think of  $A_{2,2}$  as a quantum measurement operator, and vectors of the form  $x \otimes y$  as unentangled states

(equivalently we say that vectors in this form are tensor product states, or simply “product states”). Thus the difference between  $\|A\|_{2 \rightarrow 4}^4$  and  $\|A_{2,2}\|_{2 \rightarrow 2}$  can be thought of as the extent to which the measurement  $A_{2,2}$  can notice the difference between product states and (some) entangled state.

Next, we define a matrix  $A'$  whose rows are of the form  $(x' \otimes y')^* \sqrt{A_{2,2}}$ , where  $x', y' \in \mathbb{R}^n$  range over a distribution that approximates the uniform distribution. If  $A'$  acts on a vector of the form  $x \otimes y$ , then the maximum output 4-norm (over  $L_2$ -unit vectors  $x, y$ ) is precisely  $\|A\|_{2 \rightarrow 4}$ . Intuitively, if  $A'$  acts on a highly “entangled” vector  $z$ , meaning that  $\langle z, x \otimes y \rangle$  is small for all unit vectors  $x, y$ , then  $\|A'z\|_4$  should be small. This is because  $z$  will have small overlap with  $x' \otimes y'$ , and  $A_{2,2}$  is positive semi-definite, so its off-diagonal entries can be upper-bounded in terms of its operator norm. These arguments (detailed in Section 9.2) lead to only modest bounds on  $A'$ , but then we can use an amplification argument to make the  $2 \rightarrow 4$  norm of  $A'$  depend more sensitively on that of  $A$ , at the cost of blowing up the dimension by a polynomial factor.

The reductions we achieve also permit us, in Section 9.3, to relate our Tensor-SDP algorithm with the sum-of-squares relaxation used by Doherty, Parrilo, and Spedalieri [DPS04] (henceforth DPS). We show the two relaxations are essentially equivalent, allowing us to import results proved, in some cases, with techniques from quantum information theory. One such result, from [BaCY11], requires relating  $A_{2,2}$  to a quantum measurement of the 1-LOCC form. This means that there are two  $n$ -dimensional subsystems, combined via tensor products, and  $A_{2,2}$  can be implemented as a measurement on the first subsystem followed by a measurement on the second subsystem that is chosen conditioned on the results of the first measurement. The main result of [BaCY11] proved that such LOCC measurements exhibit much better behavior under DPS, and they obtain nontrivial approximation guarantees with only  $O(\log(n)/\varepsilon^2)$  rounds. Since this is achieved by DPS, it also implies an upper bound on the error of Tensor-SDP. This upper bound is  $\varepsilon Z$ , where  $Z$  is the smallest number for which  $A_{2,2} \leq ZM$  for some 1-LOCC measurement  $M$ . While  $Z$  is not believed to be efficiently computable, it is at least  $\|A_{2,2}\|_{2 \rightarrow 2}$ , since any measurement  $M$  has  $\|M\|_{2 \rightarrow 2} \leq 1$ . To upper bound  $Z$ , we can explicitly construct  $A_{2,2}$  as a quantum measurement. This is done by the following protocol. Let  $a_1, \dots, a_m$  be the rows of  $A$ . One party performs the quantum measurement with outcomes  $\{\alpha a_i a_i^T\}_{i=1}^m$  (where  $\alpha$  is a normalization factor) and transmits the outcome  $i$  to the other party. Upon receiving message  $i$ , the second party does the two outcome measurement  $\{\beta a_i a_i^T, I - \beta a_i a_i^T\}$  and outputs 0 or 1 accordingly, where  $\beta$  is another normalization factor. The measurement  $A_{2,2}$  corresponds to the “0” outcomes. For this to be a physically realizable 1-LOCC measurement, we need  $\alpha \leq \|A^T A\|_{2 \rightarrow 2}$  and  $\beta \leq \|A\|_{2 \rightarrow \infty}^2$ . Combining these ingredients, we obtain the approximation guarantee in Theorem 2.3. More details on this argument are in Section 9.3.1.

## 4.5 Definitions and Notation

Let  $\mathcal{U}$  be some finite set. For concreteness, and without loss of generality, we can let  $\mathcal{U}$  be the set  $\{1, \dots, n\}$ , where  $n$  is some positive integer. We write  $\mathbb{E}_{\mathcal{U}} f$  to denote the average value of a function  $f: \mathcal{U} \rightarrow \mathbb{R}$  over a random point in  $\mathcal{U}$  (omitting the subscript  $\mathcal{U}$  when it is clear from the context). We let  $L_2(\mathcal{U})$  denote the space of functions  $f: \mathcal{U} \rightarrow \mathbb{R}$  endowed with the inner product  $\langle f, g \rangle = \mathbb{E}_{\mathcal{U}} fg$  and its induced norm  $\|f\| = \langle f, f \rangle^{1/2}$ . For  $p \geq 1$ , the  $p$ -norm of a function  $f \in L_2(\mathcal{U})$  is defined as  $\|f\|_p := (\mathbb{E} |f|^p)^{1/p}$ . A convexity argument shows  $\|f\|_p \leq \|f\|_q$  for  $p \leq q$ . If  $A$  is a linear operator mapping functions from  $L_2(\mathcal{U})$  to  $L_2(\mathcal{V})$ , and  $p, q \geq 1$ , then the  $p$ -to- $q$  norm of  $A$  is defined as  $\|A\|_{p \rightarrow q} = \max_{0 \neq f \in L_2(\mathcal{U})} \|Af\|_q / \|f\|_p$ . If  $V \subseteq L_2(\mathcal{U})$  is a linear subspace, then we denote



$$\|V\|_{p \rightarrow q} = \max_{f \in V} \|f\|_q / \|f\|_p.$$

**Counting norms.** In most of this paper we use *expectation norms* defined as above, but in some contexts the *counting norms* will be more convenient. We will stick to the convention that *functions* use expectation norms while *vectors* use the counting norms. For a vector  $v \in \mathbb{C}^{\mathcal{U}}$  and  $p \geq 1$ , the  $p$  *counting norm* of  $v$ , denoted  $\|v\|_p$ , is defined to be  $(\sum_{i \in \mathcal{U}} |v_i|^p)^{1/p}$ . The *counting inner product* of two vectors  $u, v \in \mathbb{R}^{\mathcal{U}}$ , denoted as  $\langle u, v \rangle$ , is defined to be  $\sum_{i \in \mathcal{U}} u_i v_i^*$ .

## 5 The Tensor-SDP algorithm

There is a very natural SoS program for the  $2 \rightarrow 4$  norm for a given linear operator  $A: L_2(\mathcal{U}) \rightarrow L_2(\mathcal{V})$ :

ALGORITHM Tensor-SDP<sup>(d)</sup>(A):

Maximize  $\tilde{\mathbb{E}}_f \|Af\|_4^4$  subject to

- $f$  is a  $d$ -f.r.v. over  $L_2(\mathcal{U})$ ,
- $\tilde{\mathbb{E}}_f (\|f\|^2 - 1)^2 = 0$ .

Note that  $\|Af\|_4^4$  is indeed a degree 4 polynomial in the variables  $\{f(u)\}_{u \in \mathcal{U}}$ . The Tensor-SDP<sup>(d)</sup> algorithm makes sense for  $d \geq 4$ , and we denote by Tensor-SDP its most basic version where  $d = 4$ . The Tensor-SDP algorithm applies not just to the  $2 \rightarrow 4$  norm, but to optimizing general polynomials over the unit ball of  $L_2(\mathcal{U})$  by replacing  $\|Af\|_4^4$  with an arbitrary polynomial  $P$ .

While we do not know the worst-case performance of the Tensor-SDP algorithm, we do know that it performs well on random instances (see Section 7), and (perhaps more relevant to the UGC) on the projector to low-degree polynomials (see Theorem 2.2). The latter is a corollary of the following result:

**Lemma 5.1.** *Over the space of  $n$ -variate Fourier polynomials<sup>9</sup>  $f$  with degree at most  $d$ ,*

$$\mathbb{E} f^4 \leq 9^d (\mathbb{E} f^2)^2,$$

where the expectations are over  $\{\pm 1\}^n$ .

*Proof.* The result is proven by a careful variant of the standard inductive proof of the hypercontractivity for low-degree polynomials (see e.g. [O'D07]). We include it in this part of the paper since it is the simplest example of how to “lift” known proofs about functions over the reals into proofs about the fictitious random variables that arise in semidefinite programming hierarchies. To strengthen the inductive hypothesis, we will prove the more general statement that for  $f$  and  $g$  being  $n$ -variate Fourier polynomials with degrees at most  $d$  and  $e$ , it holds that  $\mathbb{E} f^2 g^2 \leq 9^{\frac{d+e}{2}} (\mathbb{E} f^2) (\mathbb{E} g^2)$ . (Formally, this polynomial relation is over the linear space of pairs of  $n$ -variate Fourier polynomials  $(f, g)$ , where  $f$  has degree at most  $d$  and  $g$  has degree at most  $e$ .) The proof is by induction on the number of variables.

If one of the functions is constant (so that  $d = 0$  or  $e = 0$ ), then  $\mathbb{E} f^2 g^2 = (\mathbb{E} f^2) (\mathbb{E} g^2)$ , as desired. Otherwise, let  $f_0, f_1, g_0, g_1$  be Fourier polynomials depending only

<sup>9</sup>An  $n$ -variate Fourier polynomial with degree at most  $d$  is a function  $f: \{\pm 1\}^n \rightarrow \mathbb{R}$  of the form  $f = \sum_{\alpha \subseteq [n], |\alpha| \leq d} \hat{f}_\alpha \chi_\alpha$  where  $\chi_\alpha(x) = \prod_{i \in \alpha} x_i$ .

on  $x_1, \dots, x_{n-1}$  such that  $f(x) = f_0(x) + x_n f_1(x)$  and  $g(x) = g_0(x) + x_n g_1(x)$ . The Fourier polynomials  $f_0, f_1, g_0, g_1$  depend linearly on  $f$  and  $g$  (because  $f_0(x) = \frac{1}{2}f(x_1, \dots, x_{n-1}, 1) + \frac{1}{2}f(x_1, \dots, x_{n-1}, -1)$  and  $f_1(x) = \frac{1}{2}f(x_1, \dots, x_{n-1}, 1) - \frac{1}{2}f(x_1, \dots, x_{n-1}, -1)$ ). Furthermore, the degrees of  $f_0, f_1, g_0,$  and  $g_1$  are at most  $d, d-1, e,$  and  $e-1$ , respectively.

Since  $\mathbb{E} x_n = \mathbb{E} x_n^3 = 0$ , if we expand  $\mathbb{E} f^2 g^2 = \mathbb{E}(f_0 + x_n f_1)^2 (g_0 + x_n g_1)^2$  then the terms where  $x_n$  appears in an odd power vanish, and we obtain

$$\mathbb{E} f^2 g^2 = \mathbb{E} f_0^2 g_0^2 + f_1^2 g_1^2 + f_0^2 g_1^2 + f_1^2 g_0^2 + 4f_0 f_1 g_0 g_1$$

By expanding the square expression  $2\mathbb{E}(f_0 f_1 - g_0 g_1)^2$ , we get  $4\mathbb{E} f_0 f_1 g_0 g_1 \leq 2\mathbb{E} f_0^2 g_1^2 + f_1^2 g_0^2$  and thus

$$\mathbb{E} f^2 g^2 \leq \mathbb{E} f_0^2 g_0^2 + \mathbb{E} f_1^2 g_1^2 + 3\mathbb{E} f_0^2 g_1^2 + 3\mathbb{E} f_1^2 g_0^2. \quad (5.1)$$

Applying the induction hypothesis to all four terms on the right-hand side of (5.1) (using for the last two terms that the degree of  $f_1$  and  $g_1$  is at most  $d-1$  and  $e-1$ ),

$$\begin{aligned} \mathbb{E} f^2 g^2 &\leq 9^{\frac{d+e}{2}} (\mathbb{E} f_0^2) (\mathbb{E} g_0^2) + 9^{\frac{d+e}{2}} (\mathbb{E} f_1^2) (\mathbb{E} g_1^2) \\ &\quad + 3 \cdot 9^{\frac{d+e}{2}-1/2} (\mathbb{E} f_0^2) (\mathbb{E} g_1^2) + 3 \cdot 9^{\frac{d+e}{2}-1/2} (\mathbb{E} f_1^2) (\mathbb{E} g_0^2) \\ &= 9^{\frac{d+e}{2}} (\mathbb{E} f_0^2 + \mathbb{E} f_1^2) (\mathbb{E} g_0^2 + \mathbb{E} g_1^2). \end{aligned}$$

Since  $\mathbb{E} f_0^2 + \mathbb{E} f_1^2 = \mathbb{E}(f_0 + x_n f_1)^2 = \mathbb{E} f^2$  (using  $\mathbb{E} x_n = 0$ ) and similarly  $\mathbb{E} g_0^2 + \mathbb{E} g_1^2 = \mathbb{E} g^2$ , we derive the desired relation  $\mathbb{E} f^2 g^2 \leq 9^{\frac{d+e}{2}} (\mathbb{E} f^2) (\mathbb{E} g^2)$ .  $\square$

## 6 SoS succeeds on Unique Games integrality gaps

In this section we prove Theorem 2.6, showing that 8 rounds of the SoS hierarchy can beat the Basic SDP program on the canonical integrality gaps considered in the literature.

**Theorem 6.1** (Theorem 2.6, restated). *For sufficiently small  $\varepsilon$  and large  $k$ , and every  $n \in \mathbb{N}$ , let  $\mathcal{W}$  be an  $n$ -variable  $k$ -alphabet UNIQUE GAMES instance of the type considered in [RS09, KS09, KPS10] obtained by composing the “quotient noisy cube” instance of [KV05] with the long-code alphabet reduction of [KKMO04] so that the best assignment to  $\mathcal{W}$ ’s variable satisfies at most an  $\varepsilon$  fraction of the constraints. Then, on input  $\mathcal{W}$ , eight rounds of the SoS relaxation outputs at most  $1/100$ .*

### 6.1 Proof sketch of Theorem 6.1

The proof is very technical, as it is obtained by taking the already rather technical proofs of soundness for these instances, and “lifting” each step into the SoS hierarchy, a procedure that causes additional difficulties. The high level structure of all integrality gap instances constructed in the literature was the following: Start with a basic integrality gap instance of UNIQUE GAMES where the Basic SDP outputs  $1 - o(1)$  but the true optimum is  $o(1)$ , the alphabet size of  $\mathcal{U}$  is (necessarily)  $R = \omega(1)$ . Then, apply an *alphabet-reduction gadget* (such as the long code, or in the recent work [BGH<sup>+</sup>11] the so called “short code”) to transform  $\mathcal{U}$  into an instance  $\mathcal{W}$  with some constant alphabet size  $k$ . The soundness proof of the gadget guarantees that the true optimum of  $\mathcal{U}$  is small, while the analysis of previous works managed to “lift” the completeness proofs, and argue that the instance  $\mathcal{U}$  survives a number of rounds that tends to infinity as  $\varepsilon$  tends to zero, where  $(1 - \varepsilon)$  is the completeness

value in the gap constructions, and exact tradeoff between number of rounds and  $\varepsilon$  depends on the paper and hierarchy.

The fact that the basic instance  $\mathcal{U}$  has small integral value can be shown by appealing to hypercontractivity of low-degree polynomials, and hence can be “lifted” to the SoS world via Lemma 5.1. The bulk of the technical work is in lifting the soundness proof of the gadget. On a high level this proof involves the following components: (1) The *invariance principle* of [MOO05], saying that low influence functions cannot distinguish between the cube and the sphere; this allows us to argue that functions that perform well on the gadget must have an influential variable, and (2) the *influence decoding* procedure of [KKMO04] that maps these influential functions on each local gadget into a good global assignment for the original instance  $\mathcal{U}$ .

The invariance principle poses a special challenge, since the proof of [MOO05] uses so called “bump” functions which are not at all low-degree polynomials.<sup>10</sup> We use a weaker invariance principle, only showing that the 4 norm of a low influence function remains the same between two probability spaces that agree on the first 2 moments. Unlike the usual invariance principle, we do not move between Bernoulli variables and Gaussian space, but rather between two different distributions on the discrete cube. It turns out that for the purposes of these UNIQUE GAMES integrality gaps, the above suffices. The lifted invariance principle is proven via a “hybrid” argument similar to the argument of [MOO05], where hypercontractivity of low-degree polynomials again plays an important role.

The soundness analysis of [KKMO04] is obtained by replacing each local function with an average over its neighbors, and then choosing a random influential coordinate from the new local function as an assignment for the original uniquegames instance. We follow the same approach, though even simple tasks such as independent randomized rounding turn out to be much subtler in the lifted setting. However, it turns out that by making appropriate modification to the analysis, it can be lifted to complete the proof of Theorem 2.6.

In the following, we give a more technical description of the proof. Let  $T_{1-\eta}$  be the  $\eta$ -noise graph on  $\{\pm 1\}^R$ . Khot and Vishnoi [KV05] constructed a unique game  $\mathcal{U}$  with label-extended graph  $T_{1-\eta}$ . A solution to the level-4 SoS relaxation of  $\mathcal{U}$  is 4-f.r.v.  $h$  over  $L_2(\{\pm 1\}^R)$ . This variable satisfies  $h(x)^2 \equiv_h h(x)$  for all  $x \in \{\pm 1\}^R$  and also  $\tilde{\mathbb{E}}_h(\mathbb{E} h)^2 \leq 1/R^2$ . (The variable  $h$  encodes a 0/1 assignment to the vertices of the label-extended graph. A proper assignment assigns 1 only to a  $1/R$  fraction of these vertices.) Lemma 6.7 allows us to bound the objective value of the solution  $h$  in terms of the fourth moment  $\tilde{\mathbb{E}}_h \mathbb{E}(P_{>\lambda} h)^4$ , where  $P_{>\lambda}$  is the projector into the span of the eigenfunctions of  $T_{1-\eta}$  with eigenvalue larger than  $\lambda \approx 1/R^\eta$ . (Note that  $\mathbb{E}(P_{>\lambda} h)^4$  is a degree-4 polynomial in  $h$ .) For the graph  $T_{1-\eta}$ , we can bound the degree of  $P_{>\lambda} h$  as a Fourier polynomial (by about  $\log(R)$ ). Hence, the hypercontractivity bound (Lemma 5.1) allows us to bound the fourth moment  $\tilde{\mathbb{E}}_h \mathbb{E}(P_{>\lambda} h)^4 \leq \tilde{\mathbb{E}}_h(\mathbb{E} h^2)^2$ . By our assumptions on  $h$ , we have  $\tilde{\mathbb{E}}_h(\mathbb{E} h^2)^2 = \tilde{\mathbb{E}}_h(\mathbb{E} h)^2 \leq 1/R^2$ . Plugging these bounds into the bound of Lemma 6.7 demonstrates that the objective value of  $h$  is bounded by  $1/R^{\Omega(\eta)}$  (see Theorem 6.11).

Next, we consider a unique game  $\mathcal{W}$  obtained by composing the unique game  $\mathcal{U}$  with the alphabet reduction of [KKMO04]. Suppose that  $\mathcal{W}$  has alphabet  $\Omega = \{0, \dots, k-1\}$ . The vertex set of  $\mathcal{W}$  is  $V \times \Omega^R$  (with  $V$  being the vertex set of  $\mathcal{U}$ ). Let  $f = \{f_u\}_{u \in V}$  be a solution to the level-8 SoS relaxation of  $\mathcal{W}$ . To bound the objective value of  $f$ , we derive from it a level-4 random variable  $h$  over  $L_2(V \times [R])$ . (Encoding a function on the label-extended graph of the unique game  $\mathcal{U}$ .) We define  $h(u, r) = \text{Inf}_r^{(\leq \ell)} \bar{f}_u$ , where  $\ell \approx \log k$  and  $\bar{f}_u$  is

<sup>10</sup>A similar, though not identical, challenge arises in [BGH<sup>+</sup>11] where they need to extend the invariance principle to the “short code” setting. However, their solution does not seem to apply in our case, and we use a different approach.

a variable of  $L_2(\Omega^R)$  obtained by averaging certain values of  $f_u$  (“folding”). It is easy to show that  $h^2 \leq h$  (using [Lemma A.1](#)) and  $\tilde{\mathbb{E}}_h(\mathbb{E}h)^2 \leq \ell/R$  (bound on the total influence of low-degree Fourier polynomials). [Theorem 6.9](#) (influence decoding) allows us to bound the objective value of  $f$  in terms of the correlation of  $h$  with the label-extended graph of  $\mathcal{U}$  (in our case,  $T_{1-\eta}$ ). Here, we can use again [Theorem 6.11](#) to show that the correlation of  $h$  with the graph  $T_{1-\eta}$  is very small. (An additional challenge arises because  $h$  does not satisfy  $h^2 \equiv_h h$ , but only the weaker condition  $h^2 \leq h$ . [Corollary 6.5](#) fixes this issue by simulating independent rounding for fictitious random variables.) To prove [Theorem 6.9](#) (influence decoding), we analyze the behavior of fictitious random variables on the alphabet-reduction gadget of [\[KKMO04\]](#). This alphabet-reduction gadget essentially corresponds to the  $\varepsilon$ -noise graph  $T_{1-\varepsilon}$  on  $\Omega^R$ . Suppose  $g$  is a fictitious random variables over  $L_2(\Omega^R)$  satisfying  $g^2 \leq g$ . By [Lemma 6.7](#), we can bound the correlation of  $g$  with the graph  $T_{1-\varepsilon}$  in terms of the fourth moment of  $P_{>\lambda}g$ . At this point, the hypercontractivity bound ([Lemma 5.1](#)) is too weak to be helpful. Instead we show an “invariance principle” result ([Theorem 6.2](#)), which allows us to relate the fourth moment of  $P_{>\lambda}g$  to the fourth moment of a nicer random variable and the influences of  $g$ .

**Organization of the proof.** We now turn to the actual proof of [Theorem 6.1](#). The proof consists of lifting to the SoS hierarchy all the steps used in the analysis of previous integrality gaps, which themselves arise from hardness reductions. We start in [Section 6.2](#) by showing a sum-of-squares proof for a weaker version of [\[MOO05\]](#)’s invariance principle. Then in [Section 6.3](#) we show how one can perform independent rounding in the SoS world (this is a trivial step in proofs involving true random variables, but becomes much more subtle when dealing with SoS solutions). In [Sections 6.4](#) and [6.5](#) we lift variants of the [\[KKMO04\]](#) dictatorship test. The proof uses a SoS variant of influence decoding, which is covered in [Section 6.6](#). Together all these sections establish SoS analogs of the soundness properties of the hardness reduction used in the previous results. Then, in [Section 6.7](#) we show that analysis of the basic instance has a sum of squares proof (since it is based on hypercontractivity of low-degree polynomials). Finally in [Section 6.8](#) we combine all these tools to conclude the proof. In [Section 6.9](#) we discuss why this proof applies (with some modifications) also to the “short-code” based instances of [\[BGH<sup>+</sup>11\]](#).

## 6.2 Invariance Principle for Fourth Moment

In this section, we will give a sum-of-squares proof for a variant of the invariance principle of [\[MOO05\]](#). Instead of for general smooth functionals (usually constructed from “bump functions”), we show invariance only for the fourth moment. It turns out that invariance of the fourth moment is enough for our applications.

Let  $k = 2^t$  for  $t \in \mathbb{N}$  and let  $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_R)$  be an independent sequence<sup>11</sup> of orthonormal ensembles  $\mathcal{X}_r = (X_{r,0}, \dots, X_{r,k-1})$ . Concretely, we choose  $X_{r,i} = \chi_i(x_r)$ , where  $\chi_0, \dots, \chi_{k-1}$  is the set of characters of  $\mathbb{F}_2^t$  and  $x$  is sampled uniformly from  $(\mathbb{F}_2^t)^R$ . Every random variable over  $(\mathbb{F}_2^t)^R$  can be expressed as a multilinear polynomial over the sequence  $\mathcal{X}$ . In this sense,  $\mathcal{X}$  is maximally dependent. On the other hand, let  $\mathcal{Y} = (\mathcal{Y}_1, \dots, \mathcal{Y}_R)$  be a sequence of ensembles  $\mathcal{Y}_r = (Y_{r,0}, \dots, Y_{r,k-1})$ , where  $Y_{r,0} \equiv 1$  and  $Y_{r,j}$  are independent, unbiased  $\{\pm 1\}$  Bernoulli variables. The sequence  $\mathcal{Y}$  is maximally independent since it consists

<sup>11</sup>An *orthonormal ensemble* is a collection of orthonormal real-valued random variables, one being the constant 1. A sequence of such ensembles is *independent* if each ensemble is defined over an independent probability space. (See [\[MOO05\]](#) for details.)

of completely independent random variables.

Let  $f$  be a 4-f.r.v. over the space of multilinear polynomials with degree at most  $\ell$  and monomials indexed by  $[k]^R$ . Suppose  $\tilde{\mathbb{E}}_f \|f\|^4 \leq 1$ . (In the following, we mildly overload notation and use  $[k]$  to denote the set  $\{0, \dots, k-1\}$ .) Concretely, we can specify  $f$  by the set of monomial coefficients  $\{\hat{f}_\alpha\}_{\alpha \in [k]^R, |\alpha| \leq \ell}$ , where  $|\alpha|$  is the number of non-zero entries in  $\alpha$ . As usual, we define  $\text{Inf}_r f = \sum_{\alpha \in [k]^R, \alpha_r \neq 0} \hat{f}_\alpha^2$ . Note that  $\text{Inf}_r f$  is a degree-2 polynomial in  $f$ . (Hence, the pseudo-expectation of  $(\text{Inf}_r f)^2$  is defined.)

**Theorem 6.2** (Invariance Principle for Fourth Moment). *For  $\tau = \tilde{\mathbb{E}}_f \sum_r (\text{Inf}_r f)^2$ ,*

$$\tilde{\mathbb{E}}_f \mathbb{E}_X f^4 = \tilde{\mathbb{E}}_f \mathbb{E}_Y f^4 \pm k^{O(\ell)} \sqrt{\tau}.$$

(Since the expressions  $\mathbb{E}_X f^4$  and  $\mathbb{E}_Y f^4$  are degree-4 polynomials in  $f$ , their pseudo-expectations are defined.)

Using the SoS proof for hypercontractivity of low-degree polynomials (over the ensemble  $\mathcal{Y}$ ), the fourth moment  $\tilde{\mathbb{E}}_f \mathbb{E}_Y f^4$  is bounded in terms of the second moment  $\tilde{\mathbb{E}}_f \mathbb{E}_Y f^2$ . Since the first two moments of the ensembles  $\mathcal{X}$  and  $\mathcal{Y}$  match, we have  $\tilde{\mathbb{E}}_f \mathbb{E}_Y f^2 = \tilde{\mathbb{E}}_f \mathbb{E}_X f^2$ . Hence, we can bound the fourth moment of  $f$  over  $\mathcal{X}$  in terms of its second moment and  $\tau$ .

**Corollary 6.3.**

$$\tilde{\mathbb{E}}_f \mathbb{E}_X f^4 = 2^{O(\ell)} \tilde{\mathbb{E}}_f (\mathbb{E}_X f^2)^2 \pm k^{O(\ell)} \sqrt{\tau}.$$

(The corollary shows that for small enough  $\tau$ , the 4-norm and 2-norm of  $f$  are within a factor of  $2^{O(\ell)}$ . This bound is useful because the worst-case ratio of these norms is  $k^{O(\ell)} \gg 2^{O(\ell)}$ .)

*Proof of Theorem 6.2.* We consider the following intermediate sequences of ensembles  $\mathcal{Z}^{(r)} = (\mathcal{X}_1, \dots, \mathcal{X}_r, \mathcal{Y}_{r+1}, \dots, \mathcal{Y}_R)$ . Note that  $\mathcal{Z}^{(0)} = \mathcal{Y}$  and  $\mathcal{Z}^{(R)} = \mathcal{X}$ . For  $r \in \mathbb{N}$ , we write  $f = E_r f + D_r f$ , where  $E_r f$  is the part of  $f$  that does not depend on coordinate  $r$  and  $D_r f = f - E_r f$ . For all  $r \in \mathbb{N}$ , the following identities (between polynomials in  $f$ ) hold

$$\begin{aligned} \mathbb{E}_{\mathcal{Z}^{(r)}} f^4 - \mathbb{E}_{\mathcal{Z}^{(r-1)}} f^4 &= \mathbb{E}_{\mathcal{Z}^{(r)}} (E_r f + D_r f)^4 - \mathbb{E}_{\mathcal{Z}^{(r-1)}} (E_r f + D_r f)^4 \\ &= \mathbb{E}_{\mathcal{Z}^{(r)}} 4(E_r f)(D_r f)^3 + (D_r f)^4 - \mathbb{E}_{\mathcal{Z}^{(r-1)}} 4(E_r f)(D_r f)^3 + (D_r f)^4. \end{aligned}$$

The last step uses that the first two moments of the ensembles  $\mathcal{X}_r$  and  $\mathcal{Y}_r$  match and that  $E_r f$  does not depend on coordinate  $r$ .

Hence,

$$\mathbb{E}_X f^4 - \mathbb{E}_Y f^4 = \sum_r \mathbb{E}_{\mathcal{Z}^{(r)}} 4(E_r f)(D_r f)^3 + (D_r f)^4 - \mathbb{E}_{\mathcal{Z}^{(r-1)}} 4(E_r f)(D_r f)^3 + (D_r f)^4$$

It remains to bound the pseudo-expectation of the right-hand side. First, we consider the term  $\sum_r \mathbb{E}_{\mathcal{Z}^{(r)}} (D_r f)^4$ . The expression  $\mathbb{E}_{\mathcal{Z}^{(r)}} (D_r f)^4$  is the fourth moment of a Fourier-polynomial with degree at most  $t \cdot \ell$ . (Here, we use that the ensembles in the sequence  $\mathcal{Y}$  consist of characters of  $\mathbb{F}_2^t$ , which are Fourier polynomials of degree at most  $t$ .) Furthermore,  $\text{Inf}_r f = \mathbb{E}_{\mathcal{Z}^{(r)}} (D_r f)^2$  is the second moment of this Fourier-polynomial. Hence, by hypercontractivity of low-degree Fourier-polynomials,  $\sum_r \mathbb{E}_{\mathcal{Z}^{(r)}} (D_r f)^4 \leq \sum_r 2^{O(t \cdot \ell)} (\text{Inf}_r f)^2$ . Thus, the pseudo-expectation is at most  $\tilde{\mathbb{E}}_f \sum_r \mathbb{E}_{\mathcal{Z}^{(r)}} (D_r f)^4 \leq 2^{O(t \cdot \ell)} \tau = k^{O(\ell)} \tau$ .

Next, we consider the term  $\sum_r \mathbb{E}_{\mathcal{Z}^{(r)}}(E_r f)(D_r f)^3$ . (The remaining two terms are analogous.) To bound its pseudo-expectation, we apply Cauchy-Schwarz,

$$\tilde{\mathbb{E}}_f \sum_r \mathbb{E}_{\mathcal{Z}^{(r)}} (E_r f)(D_r f)^3 \leq \left( \tilde{\mathbb{E}}_f \sum_r \mathbb{E}_{\mathcal{Z}^{(r)}} (E_r f)^2 (D_r f)^2 \right)^{1/2} \cdot \left( \tilde{\mathbb{E}}_f \sum_r \mathbb{E}_{\mathcal{Z}^{(r)}} (D_r f)^4 \right)^{1/2} \quad (6.1)$$

Using hypercontractivity of low-degree Fourier-polynomials, we can bound the second factor of (6.1) by  $\tilde{\mathbb{E}}_f \sum_r \mathbb{E}_{\mathcal{Z}^{(r)}} (D_r f)^4 = k^{O(\ell)} \tau$ . It remains to bound the first factor of (6.1). Again by hypercontractivity,  $\mathbb{E}_{\mathcal{Z}^{(r)}} (E_r f)^2 (D_r f)^2 \leq k^{O(\ell)} \cdot \|E_r f\|^2 \cdot \|D_r f\|^2 \leq k^{O(\ell)} \|f\|^2 \cdot \|D_r f\|^2$ . By the total influence bound for low-degree polynomials, we have  $\sum_r \|D_r f\|^2 \leq \ell \|f\|^2$ . Thus  $\sum_r \mathbb{E}_{\mathcal{Z}^{(r)}} (E_r f)^2 (D_r f)^2 \leq k^{O(\ell)} \|f\|^4$ . Using the assumption  $\tilde{\mathbb{E}}_f \|f\|^4 \leq 1$ , we can bound the first factor of (6.1) by  $k^{O(\ell)}$ .

We conclude as desired that

$$\left| \tilde{\mathbb{E}}_f \mathbb{E}_X f^4 - \mathbb{E}_Y f^4 \right| \leq k^{O(\ell)} \sqrt{\tau}.$$

□

### 6.3 Interlude: Independent Rounding

In this section, we will show how to convert variables that satisfy  $f^2 \leq f$  to variables  $\bar{f}$  satisfying  $\bar{f}^2 = \bar{f}$ . The derived variables  $\bar{f}$  will inherit several properties of the original variables  $f$  (in particular, multilinear expectations). This construction corresponds to the standard independent rounding for variables with values between 0 and 1. The main challenge is that our random variables are fictitious.

Let  $f$  be a 4-f.r.v. over  $\mathbb{R}^n$ . Suppose  $f_i^2 \leq f_i$  (in terms of an unspecified jointly-distributed 4-f.r.v.). Note that for real numbers  $x$ , the condition  $x^2 \leq x$  is equivalent to  $x \in [0, 1]$ .

**Lemma 6.4.** *Let  $f$  be a 4-f.r.v. over  $\mathbb{R}^n$  and let  $i \in [n]$  such that  $f_i^2 \leq f_i$ . Then, there exists an 4-f.r.v.  $(f, \bar{f}_i)$  over  $\mathbb{R}^{n+1}$  such that  $\tilde{\mathbb{E}}_{f, \bar{f}_i} (\bar{f}_i^2 - \bar{f}_i)^2 = 0$  and for every polynomial  $P$  which is linear in  $\bar{f}_i$  and has degree at most 4,*

$$\tilde{\mathbb{E}}_{f, \bar{f}_i} P(f, \bar{f}_i) = \tilde{\mathbb{E}}_f P(f, f_i).$$

*Proof.* We define the pseudo-expectation functional  $\tilde{\mathbb{E}}_{f, \bar{f}_i}$  as follows: For every polynomial  $P$  in  $(f, \bar{f}_i)$  of degree at most 4, let  $P'$  be the polynomial obtained by replacing  $\bar{f}_i^2$  by  $\bar{f}_i$  until  $P'$  is (at most) linear in  $\bar{f}_i$ . (In other words, we reduce  $P$  modulo the relation  $\bar{f}_i^2 = \bar{f}_i$ .) We define  $\tilde{\mathbb{E}}_{f, \bar{f}_i} P(f, \bar{f}_i) = \tilde{\mathbb{E}}_f P'(f, f_i)$ . With this definition,  $\tilde{\mathbb{E}}_{f, \bar{f}_i} (\bar{f}_i^2 - \bar{f}_i)^2 = 0$ . The operator  $\tilde{\mathbb{E}}_{f, \bar{f}_i}$  is clearly linear (since  $(P + Q)' = P' + Q'$ ). It remains to verify positivity. Let  $P$  be a polynomial of degree at most 4. We will show  $\mathbb{E}_{f, \bar{f}_i} P^2(f, \bar{f}_i) \geq 0$ . Without loss of generality  $P$  is linear in  $\bar{f}_i$ . We express  $P = Q + \bar{f}_i R$ , where  $Q$  and  $R$  are polynomials in  $f$ . Then,  $(P^2)' = Q^2 + 2\bar{f}_i QR + \bar{f}_i^2 R^2$ . Using our assumption  $f_i^2 \leq f_i$ , we get  $(P^2)'(f, f_i) = Q^2 + 2f_i QR + f_i R^2 \geq Q^2 + 2f_i QR + f_i^2 R^2 = P^2(f, f_i)$ . It follows as desired that

$$\tilde{\mathbb{E}}_{f, \bar{f}_i} P^2 = \tilde{\mathbb{E}}_f (P^2)'(f, f_i) \geq \tilde{\mathbb{E}}_f P^2(f, f_i) \geq 0.$$

□

**Corollary 6.5.** *Let  $f$  be a 4-f.r.v. over  $\mathbb{R}^n$  and let  $I \subseteq [n]$  such that  $f_i^2 \leq f_i$  for all  $i \in I$ . Then, there exists an 4-f.r.v.  $(f, \bar{f}_I)$  over  $\mathbb{R}^{n+|I|}$  such that  $\tilde{\mathbb{E}}_{f, \bar{f}_I} (f_i^2 - \bar{f}_i)^2 = 0$  for all  $i \in I$  and for every polynomial  $P$  which is multilinear in the variables  $\{f_i\}_{i \in I}$  and has degree at most 4,*

$$\tilde{\mathbb{E}}_{f, \bar{f}_I} P(f, \bar{f}_I) = \tilde{\mathbb{E}}_f P(f, f_I).$$

#### 6.4 Dictatorship Test for Small-Set Expansion

Let  $\Omega = \{0, \dots, k-1\}$  and let  $T_{1-\varepsilon}$  be the noise graph on  $\Omega^R$  with second largest eigenvalue  $1 - \varepsilon$ . Let  $f$  be a 4-f.r.v. over  $L_2(\Omega^R)$ . Suppose  $f^2 \leq f$  (in terms of an unspecified jointly-distributed 4-f.r.v.). Note that for real numbers  $x$ , the condition  $x^2 \leq x$  is equivalent to  $x \in [0, 1]$ .

The following theorem is an analog of the ‘‘Majority is Stablest’’ result [MOO05].

**Theorem 6.6.** *Suppose  $\tilde{\mathbb{E}}_f (\mathbb{E} f)^2 \leq \delta^2$ . Let  $\tau = \tilde{\mathbb{E}}_f \sum_r (\text{Inf}_r^{(\leq \ell)} f)^2$  for  $\ell = \Omega(\log(1/\delta))$ . Then,*

$$\tilde{\mathbb{E}}_f \langle f, T_{1-\varepsilon} f \rangle \leq \delta^{1+\Omega(\varepsilon)} + k^{O(\log(1/\delta))} \cdot \tau^{1/8}.$$

(Here, we assume that  $\varepsilon$ ,  $\delta$  and  $\tau$  are sufficiently small.)

The previous theorem is about graph expansion (measured by the quadratic form  $\langle f, T_{1-\varepsilon} f \rangle$ ). The following lemma allows us to relate graph expansion to the 4-norm of the projection of  $f$  into the span of the eigenfunctions of  $T_{1-\varepsilon}$  with significant eigenvalue. We will be able to bound this 4-norm in terms of the influences of  $f$  (using the invariance principle in the previous section).

**Lemma 6.7.** *Let  $f$  be a 4-f.r.v. over  $L_2(\Omega^R)$ . Suppose  $f^2 \leq f$  (in terms of unspecified jointly-distributed 4-f.r.v.  $s$ ). Then for all  $\lambda > 0$ ,*

$$\tilde{\mathbb{E}}_f \langle f, T_{1-\varepsilon} f \rangle \leq \left( \tilde{\mathbb{E}}_f \mathbb{E} f \right)^{3/4} \left( \tilde{\mathbb{E}}_f \mathbb{E} (P_{>\lambda} f)^4 \right)^{1/4} + \lambda \tilde{\mathbb{E}}_f \mathbb{E} f.$$

Here,  $P_{>\lambda}$  is the projector into the span of the eigenfunctions of  $T_{1-\varepsilon}$  with eigenvalue larger than  $\lambda$ .

*Proof.* The following relation between polynomials holds

$$\langle f, T_{1-\varepsilon} f \rangle \leq \mathbb{E} f \cdot (P_{>\lambda} f) + \lambda \mathbb{E} f^2.$$

By [Corollary 6.5](#), there exists a 4-f.r.v.  $(f, \bar{f})$  over  $L_2(\Omega^R) \times L_2(\Omega^R)$  such that  $\bar{f}^2 \equiv_{\bar{f}} \bar{f}$ . Then,

$$\begin{aligned} \tilde{\mathbb{E}}_f \mathbb{E} f \cdot (P_{>\lambda} f) &= \tilde{\mathbb{E}}_{f, \bar{f}} \mathbb{E} \bar{f} \cdot (P_{>\lambda} f) \quad (\text{using linearity in } \bar{f}) \\ &= \tilde{\mathbb{E}}_{f, \bar{f}} \mathbb{E} \bar{f}^3 \cdot (P_{>\lambda} f) \quad (\text{using } \bar{f}^2 \equiv_{\bar{f}} \bar{f}) \\ &\leq \left( \tilde{\mathbb{E}}_{\bar{f}} \mathbb{E} \bar{f}^4 \right)^{3/4} \cdot \left( \tilde{\mathbb{E}}_f \mathbb{E} (P_{>\lambda} f)^4 \right)^{1/4} \quad (\text{using Lemma A.5 (H\"older)}) \\ &= \left( \tilde{\mathbb{E}}_{\bar{f}} \mathbb{E} \bar{f} \right)^{3/4} \cdot \left( \tilde{\mathbb{E}}_f \mathbb{E} (P_{>\lambda} f)^4 \right)^{1/4} \quad (\text{using } \bar{f}^2 \equiv_{\bar{f}} \bar{f}) \\ &= \left( \tilde{\mathbb{E}}_f \mathbb{E} f \right)^{3/4} \cdot \left( \tilde{\mathbb{E}}_f \mathbb{E} (P_{>\lambda} f)^4 \right)^{1/4} \quad (\text{using linearity in } \bar{f}) \quad \square \end{aligned}$$

*Proof of Theorem 6.6.* By Lemma 6.7,

$$\tilde{\mathbb{E}}_f \langle f, T_{1-\varepsilon} f \rangle \leq (\tilde{\mathbb{E}}_f \mathbb{E} f)^{3/4} (\tilde{\mathbb{E}}_f \mathbb{E} (P_{>\lambda} f)^4)^{1/4} + \lambda \tilde{\mathbb{E}}_f \mathbb{E} f^2.$$

Using Corollary 6.3,

$$\tilde{\mathbb{E}}_f \langle f, T_{1-\varepsilon} f \rangle \leq 2^{O(\ell)} \cdot (\tilde{\mathbb{E}}_f \mathbb{E} f)^{3/4} (\tilde{\mathbb{E}}_f \mathbb{E} f^2)^{1/4} + \sqrt{\tau} \cdot k^{O(\ell)} + \lambda \tilde{\mathbb{E}}_f \mathbb{E} f^2.$$

Here,  $\ell = \log(1/\lambda)/\varepsilon$ . Using the relation  $f^2 \leq f$  and our assumption  $\tilde{\mathbb{E}}_f (\mathbb{E} f)^2 \leq \delta^2$ , we get  $\tilde{\mathbb{E}}_f \mathbb{E} f^2 \leq \tilde{\mathbb{E}}_f \mathbb{E} f \leq (\tilde{\mathbb{E}}_f \mathbb{E} f^2)^{1/2} \leq \delta$  (by Cauchy–Schwarz). Hence,

$$\begin{aligned} \tilde{\mathbb{E}}_f \langle f, T_{1-\varepsilon} f \rangle &\leq (1/\lambda)^{O(1/\varepsilon)} \delta^{3/4} (\delta^2 + \sqrt{\tau} \cdot (1/\lambda)^{O(\log k)/\varepsilon})^{1/4} + \lambda \delta \\ &\leq (1/\lambda)^{O(1/\varepsilon)} \delta^{5/4} + (1/\lambda)^{O(\log k)/\varepsilon} \delta^{3/4} \tau^{1/8} + \lambda \cdot \delta. \end{aligned}$$

To balance the terms  $(1/\lambda)^{O(1/\varepsilon)} \delta^{5/4}$  and  $\lambda \delta$ , we choose  $\lambda = \delta^{\Omega(\varepsilon)}$ . We conclude the desired bound,

$$\tilde{\mathbb{E}}_f \langle f, T_{1-\varepsilon} f \rangle \leq \delta^{1+\Omega(\varepsilon)} + k^{O(\log(1/\delta))} \cdot \tau^{1/8}. \quad \square$$

## 6.5 Dictatorship Test for Unique Games

Let  $\Omega = \mathbb{Z}_k$  (cyclic group of order  $k$ ) and let  $f$  be a 4-f.r.v. over  $L_2(\Omega \times \Omega^R)$ . Here,  $f(a, x)$  is intended to be 0/1 variable indicating whether symbol  $a$  is assigned to the point  $x$ .

The following graph  $T'_{1-\varepsilon}$  on  $\Omega \times \Omega^R$  corresponds to the 2-query dictatorship test for UNIQUE GAMES [KKMO04],

$$T'_{1-\varepsilon} f(a, x) = \mathbb{E}_{c \in \Omega} \mathbb{E}_{y \sim_{1-\varepsilon} x} f(a + c, y - c \cdot \mathbf{I}).$$

Here,  $y \sim_{1-\varepsilon} x$  means that  $y$  is a random neighbor of  $x$  in the graph  $T_{1-\varepsilon}$  (the  $\varepsilon$ -noise graph on  $\Omega^R$ ).

We define  $\bar{f}(x) := \mathbb{E}_{c \in \Omega} f(c, x - c \cdot \mathbf{I})$ . (We think of  $\bar{f}$  as a variable over  $L_2(\Omega^R)$ .) Then, the following polynomial identity (in  $f$ ) holds

$$\langle f, T'_{1-\varepsilon} f \rangle = \langle \bar{f}, T_{1-\varepsilon} \bar{f} \rangle.$$

**Theorem 6.8.** *Suppose  $f^2 \leq f$  and  $\tilde{\mathbb{E}}_f (\mathbb{E} f)^2 \leq \delta^2$ . Let  $\tau = \tilde{\mathbb{E}}_f \sum_r (\text{Inf}_r^{(\leq \ell)} \bar{f})^2$  for  $\ell = \Omega(\log(1/\delta))$ . Then,*

$$\tilde{\mathbb{E}}_f \langle f, T'_{1-\varepsilon} f \rangle \leq \delta^{1+\Omega(\varepsilon)} + k^{O(\log(1/\delta))} \cdot \tau^{1/8}.$$

(Here, we assume that  $\varepsilon$ ,  $\delta$  and  $\tau$  are sufficiently small.)

*Proof.* Apply Theorem 6.6 to bound  $\tilde{\mathbb{E}}_f \langle \bar{f}, T_{1-\varepsilon} \bar{f} \rangle$ . Use that fact that  $\mathbb{E} f = \mathbb{E} \bar{f}$  (as polynomials in  $f$ ).  $\square$

## 6.6 Influence Decoding

Let  $\mathcal{U}$  be a unique game with vertex set  $V$  and alphabet  $[R]$ . Recall that we represent  $\mathcal{U}$  as a distribution over triples  $(u, v, \pi)$  where  $u, v \in V$  and  $\pi$  is a permutation of  $[R]$ . The triples encode the constraints of  $\mathcal{U}$ . We assume that the unique game  $\mathcal{U}$  is regular in the same that every vertex participates in the same fraction of constraints.



Let  $\Omega = \mathbb{Z}_k$  (cyclic group of order  $k$ ). We reduce  $\mathcal{U}$  to a unique game  $\mathcal{W} = \mathcal{W}_{\varepsilon,k}(\mathcal{U})$  with vertex set  $V \times \Omega^R$  and alphabet  $\Omega$ . Let  $f = \{f_u\}_{u \in V}$  be a variable over  $L_2(\Omega \times \Omega^R)^V$ . The unique game  $\mathcal{W}$  corresponds to the following quadratic form in  $f$ ,

$$\langle f, \mathcal{W}f \rangle := \mathbb{E}_{u \in V} \mathbb{E}_{\substack{(u,v,\pi) \sim \mathcal{U}|u \\ (u,v',\pi') \sim \mathcal{U}|u}} \langle f_v^{(\pi)}, T'_{1-\varepsilon} f_{v'}^{(\pi')} \rangle.$$

Here,  $(u, v, \pi) \sim \mathcal{U} | u$  denotes a random constraint of  $\mathcal{U}$  incident to vertex  $u$ , the graph  $T'_{1-\varepsilon}$  corresponds to the dictatorship test of UNIQUE GAMES defined in Section 6.5, and  $f_v^{(\pi)}(a, x) = f_v(a, \pi.x)$  is the function obtained by permuting the last  $R$  coordinates according to  $\pi$  (where  $\pi.x(i) = x_{\pi(i)}$ ).

We define  $g_u = \mathbb{E}_{(u,v,\pi) \sim \mathcal{U}|u} f_v^{(\pi)}$ . Then,

$$\langle f, \mathcal{W}f \rangle = \mathbb{E}_{u \in V} \langle g_u, T'_{1-\varepsilon} g_u \rangle. \quad (6.2)$$

**Bounding the value of SoS solutions.** Let  $f = \{f_u\}_{u \in V}$  be a solution to the level- $d$  SoS relaxation for the unique game  $\mathcal{W}$ . In particular,  $f$  is a  $d$ -f.r.v. over  $L_2(\Omega \times \Omega^R)^V$ . Furthermore,  $\tilde{\mathbb{E}}_f(\mathbb{E} f_u)^2 \leq 1/k^2$  for all vertices  $u \in V$ .

By applying Theorem 6.8 to (6.2), we can bound the objective value of  $f$

$$\tilde{\mathbb{E}}_f \langle f, \mathcal{W}f \rangle \leq 1/k^{1+\Omega(\varepsilon)} + k^{O(\log k)} \left( \tilde{\mathbb{E}}_f \mathbb{E}_{u \in V} \tau_u \right)^{1/8},$$

where  $\tau_u = \sum_r (\text{Inf}_r^{(\leq \ell)} \bar{g}_u)^2$ ,  $\bar{g}_u(x) = \mathbb{E}_{(u,v,\pi) \sim \mathcal{U}|u} \bar{f}_v^{(\pi)}$ , and  $\bar{f}_v(x) = \mathbb{E}_{c \in \Omega} f_v(c, x - c \cdot \mathbf{I})$ .

Since  $\text{Inf}_r^{(\leq \ell)}$  is a positive semidefinite form,

$$\tau_u \leq \sum_r \left( \mathbb{E}_{(u,v,\pi) \sim \mathcal{U}|u} \text{Inf}_r^{(\leq \ell)} \bar{f}_v^{(\pi)} \right)^2 = \sum_r \left( \mathbb{E}_{(u,v,\pi) \sim \mathcal{U}|u} \text{Inf}_{\pi(r)}^{(\leq \ell)} \bar{f}_v \right)^2.$$

Let  $h$  be the level- $d/2$  fictitious random variable over  $L_2(V \times [R])$  with  $h(u, r) = \text{Inf}_r^{(\leq \ell)} \bar{f}_u$ . Let  $G_{\mathcal{U}}$  be the label-extended graph of the unique game  $\mathcal{U}$ . Then, the previous bound on  $\tau_u$  shows that  $\mathbb{E}_{u \in V} \tau_u \leq R \cdot \|G_{\mathcal{U}} h\|^2$ . Lemma A.1 shows that  $h^2 \leq h$ . On the other hand,  $\sum_r h(u, r) \leq \ell \|\bar{f}_u\|^2 \leq \ell \|f_u\|^2$  (bound on the total influence of low-degree Fourier polynomials). In particular,  $\mathbb{E} h \leq \ell \mathbb{E}_{u \in V} \|f_u\|^2 / R$ . Since  $f$  is a valid SoS solution for the unique game  $\mathcal{W}$ , we have  $\tilde{\mathbb{E}}_f \|f_u\|^d \leq 1/k^{d/2}$  for all  $u \in V$ . (Here, we assume that  $d$  is even.) It follows that  $\tilde{\mathbb{E}}_h (\mathbb{E} h)^{d/2} \leq (\frac{\ell}{kR})^{d/2}$ .

The arguments in this subsection imply the following theorem.

**Theorem 6.9.** *The optimal value of the level- $d$  SoS relaxation for the unique game  $\mathcal{W} = \mathcal{W}_{\varepsilon,k}(\mathcal{U})$  is bounded from above by*

$$1/k^{\Omega(\varepsilon)} + k^{O(\log k)} \left( R \cdot \max_h \tilde{\mathbb{E}} \|G_{\mathcal{U}} h\|^2 \right)^{1/8},$$

where the maximum is over all level- $d/2$  fictitious random variables  $h$  over  $L_2(V \times [R])$  satisfying  $h^2 \leq h$  and  $\tilde{\mathbb{E}}_h (\mathbb{E} h)^{d/2} \leq \ell/R^{d/2}$ .

**Remark 6.10.** Since the quadratic form  $\|G_{\mathcal{U}} h\|^2$  has only nonnegative coefficients (in the standard basis), we can use Corollary 6.5 to ensure that the level- $d/2$  random variable  $h$  satisfies in addition  $h^2 \equiv_h h$ .

## 6.7 Certifying Small-Set Expansion

Let  $T_{1-\varepsilon}$  be a the noise graph on  $\{\pm 1\}^R$  with second largest eigenvalue  $1 - \varepsilon$ .

**Theorem 6.11.** *Let  $f$  be level-4 fictitious random variables over  $L_2(\{\pm 1\}^R)$ . Suppose that  $f^2 \leq f$  (in terms of unspecified jointly-distributed level-4 fictitious random variables) and that  $\tilde{\mathbb{E}}_f(\mathbb{E} f)^2 \leq \delta^2$ . Then,*

$$\tilde{\mathbb{E}}_f \langle f, T_{1-\varepsilon} f \rangle \leq \delta^{1+\Omega(\varepsilon)}.$$

*Proof.* By [Lemma 6.7](#) (applying it for the case  $\Omega = \{0, 1\}$ ), for every  $\lambda > 0$ ,

$$\tilde{\mathbb{E}}_f \langle f, T_{1-\varepsilon} f \rangle \leq (\tilde{\mathbb{E}}_f \mathbb{E} f)^{3/4} (\tilde{\mathbb{E}}_f \mathbb{E}(P_{>\lambda} f)^4)^{1/4} + \lambda \tilde{\mathbb{E}}_f \mathbb{E} f.$$

For the graph  $T_{1-\varepsilon}$ , the eigenfunctions with eigenvalue larger than  $\lambda$  are characters with degree at most  $\log(1/\lambda)/\varepsilon$ . Hence, [Lemma 5.1](#) implies  $\mathbb{E}(P_{>\lambda} f)^4 \leq (1/\lambda)^{O(1/\varepsilon)} \|f\|^4$ . Since  $f^2 \leq f$ , we have  $\|f\|^4 \leq (\mathbb{E} f)^2$ . Hence,  $\tilde{\mathbb{E}}_f \mathbb{E}(P_{>\lambda} f)^4 \leq (1/\lambda)^{O(1/\varepsilon)} \delta^2$ . Plugging in, we get

$$\tilde{\mathbb{E}}_f \langle f, T_{1-\varepsilon} f \rangle \leq (1/\lambda)^{O(1/\varepsilon)} \delta^{5/4} + \lambda \cdot \delta.$$

To balance the terms, we choose  $\lambda = \delta^{\Omega(\varepsilon)}$ , which gives the desired bound.  $\square$

## 6.8 Putting Things Together

Let  $T_{1-\eta}$  be a the noise graph on  $\{\pm 1\}^R$  with second largest eigenvalue  $1-\eta$ . Let  $\mathcal{U} = \mathcal{U}_{\eta,R}$  be an instance of UNIQUE GAMES with label-extended graph  $G_{\mathcal{U}} = T_{1-\eta}$  (e.g., the construction in [\[KV05\]](#)).

Combining [Theorem 6.9](#) (with  $d = 4$ ) and [Theorem 6.11](#) gives the following result.

**Theorem 6.12.** *The optimal value of the level-8 SoS relaxation for the unique game  $\mathcal{W} = \mathcal{W}_{\varepsilon,k}(\mathcal{U}_{\eta,R})$  is bounded from above by*

$$1/k^{\Omega(\varepsilon)} + k^{O(\log k)} \cdot R^{-\Omega(\eta)}.$$

*In particular, the optimal value of the relaxation is close to  $1/k^{\Omega(\varepsilon)}$  if  $\log R \gg (\log k)^2/\eta$ .*

## 6.9 Refuting Instances based on Short Code

Let  $\mathcal{U}' = \mathcal{U}'_{\eta,R}$  be an instance of UNIQUE GAMES according to the basic construction in [\[BGH<sup>+</sup>11\]](#). (The label-extended graph of  $\mathcal{U}'$  will be a subgraph of  $T_{1-\varepsilon}$  induced by the subset of  $\{\pm 1\}^R$  corresponding to a Reed–Muller code, that is, evaluations of low-degree  $\mathbb{F}_2$ -polynomials.)

Let  $\mathcal{W}' = \mathcal{W}'_{\varepsilon,k}(\mathcal{U}'_{\eta,R})$  be the unique game obtained by applying the short-code alphabet reduction of [\[BGH<sup>+</sup>11\]](#).

The following analog of [Theorem 6.12](#) holds.

**Theorem 6.13.** *The optimal value of the level-8 SoS relaxation for the unique game  $\mathcal{W}' = \mathcal{W}'_{\varepsilon,k}(\mathcal{U}'_{\eta,R})$  is bounded from above by*

$$1/k^{\Omega(\varepsilon)} + k^{O(\log k)} \cdot R^{-\Omega(\eta)}.$$

*In particular, the optimal value of the relaxation is close to  $1/k^{\Omega(\varepsilon)}$  if  $\log R \gg (\log k)^2/\eta$ .*

The proof of [Theorem 6.13](#) is almost literally the same as the proof of [Theorem 6.12](#). In the following, we sketch the main arguments why the proof doesn't have to change. First, several of the results of the previous sections apply to general graphs and instances of `UNIQUE GAMES`. In particular, [Lemma 6.7](#) applies to general graphs and [Theorem 6.9](#) applies to general gadget-composed instances of unique games assuming a “Majority is Stablest” result for the gadget. In fact, the only parts that require further justification are the invariance principle ([Theorem 6.2](#)) and hypercontractivity bound ([Lemma 5.1](#)). Both the invariance principle and the hypercontractivity bound are about the fourth moment of a low-degree Fourier polynomial (whose coefficients are fictitious random variables). For the construction of [\[BGH<sup>+</sup>11\]](#), we need to argue about the fourth moment with respect to a different distribution over inputs. (Instead of the uniform distribution, [\[BGH<sup>+</sup>11\]](#) considers a distribution over inputs related to the Reed–Muller code.) However, this distribution happens to be  $k$ -wise independent for  $k/4$  larger than the degree of our Fourier polynomial. Hence, as a degree-4 polynomial in Fourier coefficients, the fourth moment with respect to the [\[BGH<sup>+</sup>11\]](#)-input distribution is the same as with respect to the uniform distribution, which considered here.

## 7 Hypercontractivity of random operators

We already saw that the Tensor-SDP algorithm provides non-trivial guarantees on the  $2 \rightarrow 4$  norms of the projector to low-degree polynomials. In this section we show that it also works for a natural but very different class of instances, namely random linear operators.

Let  $A = \sum_{i=1}^m e_i a_i^T / \sqrt{n}$ , where  $e_i$  is the vector with a 1 in the  $i^{\text{th}}$  position, and each  $a_i$  is chosen i.i.d. from a distribution  $\mathcal{D}$  on  $\mathbb{R}^n$ . Three natural possibilities are

1.  $\mathcal{D}_{\text{sign}}$ : the uniform distribution over  $\{-1, 1\}^n$
2.  $\mathcal{D}_{\text{Gaussian}}$ : a vector of  $n$  independent Gaussians with mean zero and variance 1
3.  $\mathcal{D}_{\text{unit}}$ : a uniformly random unit vector on  $\mathbb{R}^n$ .

Our arguments will apply to any of these cases, or even to more general nearly-unit vectors with bounded sub-Gaussian moment (details below).

Before discussing the performance of Tensor-SDP, we will discuss how the  $2 \rightarrow 4$ -norm of  $A$  behaves as a function of  $n$  and  $m$ . We can gain intuition by considering two limits in the case of  $\mathcal{D}_{\text{Gaussian}}$ . If  $n = 1$ , then  $\|A\|_{2 \rightarrow 4} = \|a\|_4$ , for a random Gaussian vector  $a$ . For large  $m$ ,  $\|a\|_4$  is likely to be close to  $3^{1/4}$ , which is the fourth moment of a mean-zero unit-variance Gaussian. By Dvoretzky's theorem [\[Pis99\]](#), this behavior can be shown to extend to higher values of  $n$ . Indeed, there is a universal  $c > 0$  such that if  $n \leq c \sqrt{m} \varepsilon^2$ , then w.h.p.  $\|A\|_{2 \rightarrow 4} \leq 3^{1/4} + \varepsilon$ . In this case, the maximum value of  $\|Ax\|_4$  looks roughly the same as the average or the minimum value, and we also have  $\|Ax\|_4 \geq (3^{1/4} - \varepsilon)\|x\|_2$  for all  $x \in \mathbb{R}^n$ . In the cases of  $\mathcal{D}_{\text{sign}}$  and  $\mathcal{D}_{\text{unit}}$ , the situation is somewhat more complicated, but for large  $n$ , their behavior becomes similar to the Gaussian case.

On the other hand a simple argument (a variant of [Corollary 10.2](#)) shows that  $\|A\|_{2 \rightarrow 4} \geq n^{1/2}/m^{1/4}$  for any (not only random)  $m \times n$  matrix with all  $\pm 1/\sqrt{n}$  entries. A nearly identical bound applies for the case when the  $a_i$  are arbitrary unit or near-unit vectors. Thus, in the regime where  $n \geq \omega(\sqrt{m})$ , we always have  $\|A\|_{2 \rightarrow 4} \geq \omega(1)$ .

The following theorem shows that Tensor-SDP achieves approximately the correct answer in both regimes.

**Theorem 7.1.** Let  $a_1, \dots, a_m$  be drawn i.i.d. from a distribution  $\mathcal{D}$  on  $\mathbb{R}^n$  with  $\mathcal{D} \in \{\mathcal{D}_{\text{Gaussian}}, \mathcal{D}_{\text{sign}}, \mathcal{D}_{\text{unit}}\}$ , and let  $A = \sum_{i=1}^m e_i a_i^T / \sqrt{n}$ . Then w.h.p.  $\text{Tensor-SDP}(A) \leq 3 + c \max(\frac{n}{\sqrt{m}}, \frac{n^2}{m})$  for some constant  $c > 0$ .

From Theorem 7.1 and the fact that  $\|A\|_{2 \rightarrow 4}^4 \leq \text{Tensor-SDP}(A)$ , we obtain:

**Corollary 7.2.** Let  $A$  be as in Theorem 7.1. Then  $\exists c > 0$  such that w.h.p.

$$\|A\|_{2 \rightarrow 4} \leq \begin{cases} 3^{1/4} + c \frac{n}{\sqrt{m}} & \text{if } n \leq \sqrt{m} \\ c \frac{n^{1/2}}{m^{1/4}} & \text{if } n > \sqrt{m} \end{cases} \quad (7.1)$$

Before proving Theorem 7.1, we introduce some more notation. This will in fact imply that Theorem 7.1 applies to a broader class of distributions. For a distribution  $\mathcal{D}$  on  $\mathbb{R}^N$ , define the  $\psi_p$  norm  $\|\mathcal{D}\|_{\psi_p}$  to be the smallest  $C > 0$  such that

$$\max_{v \in S(\mathbb{R}^N)} \mathbb{E}_{a \sim \mathcal{D}} e^{\frac{\langle v, a \rangle^p N^{p/2}}{C^p}} \leq 2, \quad (7.2)$$

or  $\infty$  if no finite such  $C$  exists. We depart from the normal convention by including a factor of  $N^{p/2}$  in the definition, to match the scale of [ALPTJ11]. The  $\psi_2$  norm (technically a seminorm) is also called the sub-Gaussian norm of the distribution. One can verify that for each of the above examples (sign, unit and Gaussian vectors),  $\psi_2(\mathcal{D}) \leq O(1)$ .

We also require that  $\mathcal{D}$  satisfies a *boundedness* condition with constant  $K \geq 1$ , defined as

$$\mathbb{P} \left[ \max_{i \in [m]} \|a_i\|_2 > K \max(1, (m/N)^{1/4}) \right] \leq e^{-\sqrt{N}}. \quad (7.3)$$

Similarly,  $K$  can be taken to be  $O(1)$  in each case that we consider.

We will require a following result of [ALPTJ10, ALPTJ11] about the convergence of sums of i.i.d rank-one matrices.

**Lemma 7.3** ([ALPTJ11]). Let  $\mathcal{D}'$  be a distribution on  $\mathbb{R}^N$  such that  $\mathbb{E}_{v \sim \mathcal{D}'} v v^T = I$ ,  $\|\mathcal{D}'\|_{\psi_1} \leq \psi$  and (7.3) holds for  $\mathcal{D}'$  with constant  $K$ . Let  $v_1, \dots, v_m$  be drawn i.i.d. from  $\mathcal{D}'$ . Then with probability  $\geq 1 - 2 \exp(-c \sqrt{N})$ , we have

$$(1 - \varepsilon)I \leq \frac{1}{m} \sum_{i=1}^m v_i v_i^T \leq (1 + \varepsilon)I, \quad (7.4)$$

where  $\varepsilon = C(\psi + K)^2 \max(N/m, \sqrt{N/m})$  with  $c, C > 0$  universal constants.

The  $N \leq m$  case (when the  $\sqrt{N/m}$  term is applicable) was proven in Theorem 1 of [ALPTJ11], and the  $N > m$  case (i.e. when the max is achieved by  $N/m$ ) was proven in Theorem 2 of [ALPTJ11] (see also Theorem 3.13 of [ALPTJ10]).

*Proof of Theorem 7.1.* Define  $A_{2,2} = \frac{1}{m} \sum_{i=1}^m a_i a_i^T \otimes a_i a_i^T$ . For  $n^2 \times n^2$  real matrices  $X, Y$ , define  $\langle X, Y \rangle := \text{Tr } X^T Y / n^2 = \mathbb{E}_{i,j \in [n]} X_{i,j} Y_{i,j}$ . Additionally define the convex set  $\mathcal{X}$  to be the set of  $n^2 \times n^2$  real matrices  $X = (X_{(i_1, i_2), (i_3, i_4)})_{i_1, i_2, i_3, i_4 \in [n]}$  with  $X \geq 0$ ,  $\mathbb{E}_{i,j \in [n]} X_{(i,j), (i,j)} = 1$  and  $X_{(i_1, i_2), (i_3, i_4)} = X_{(i_{\pi(1)}, i_{\pi(2)}), (i_{\pi(3)}, i_{\pi(4)})}$  for any permutation  $\pi \in \mathcal{S}_4$ . Finally, let  $h_{\mathcal{X}}(Y) := \max_{X \in \mathcal{X}} \langle X, Y \rangle$ . It is straightforward to show (c.f. Lemma 9.3) that

$$\text{Tensor-SDP}(A) = h_{\mathcal{X}}(A_{2,2}) = \max_{X \in \mathcal{X}} \langle X, A_{2,2} \rangle. \quad (7.5)$$

We note that if  $\mathcal{X}$  were defined without the symmetry constraint, it would simply be the convex hull of  $xx^T$  for unit vectors  $x \in \mathbb{R}^{n^2}$  and  $\text{Tensor-SDP}(A)$  would simply be the largest eigenvalue of  $A_{2,2}$ . However, we will later see that the symmetry constraint is crucial to  $\text{Tensor-SDP}(A)$  being  $O(1)$ .

Our strategy will be to analyze  $A_{2,2}$  by applying Lemma 7.3 to the vectors  $v_i := \Sigma^{-1/2}(a_i \otimes a_i)$ , where  $\Sigma = \mathbb{E} a_i a_i^T \otimes a_i a_i^T$ , and  $^{-1/2}$  denotes the pseudo-inverse. First, observe that, just as the  $\psi_2$  norm of the distribution over  $a_i$  is constant, a similar calculation can verify that the  $\psi_1$  norm of the distribution over  $a_i \otimes a_i$  is also constant. Next, we have to argue that  $\Sigma^{-1/2}$  does not increase the norm by too much.

To do so, we compute  $\Sigma$  for each distribution over  $a_i$  that we have considered. Let  $F$  be the operator satisfying  $F(x \otimes y) = y \otimes x$  for any  $x, y \in \mathbb{R}^n$ ; explicitly  $F = P_n((1, 2))$  from (9.9). Define

$$\Phi := \sum_{i=1}^n e_i \otimes e_i \quad (7.6)$$

$$\Delta := \sum_{i=1}^n e_i e_i^T \otimes e_i e_i^T \quad (7.7)$$

Direct calculations (omitted) can verify that the cases of random Gaussian vectors, random unit vectors and random  $\pm 1$  vectors yield respectively

$$\Sigma_{\text{Gaussian}} = I + F + \Phi\Phi^T \quad (7.8a)$$

$$\Sigma_{\text{unit}} = \frac{n}{n+1} \Sigma_{\text{Gaussian}} \quad (7.8b)$$

$$\Sigma_{\text{sign}} = \Sigma_{\text{Gaussian}} - 2\Delta \quad (7.8c)$$

In each case, the smallest nonzero eigenvalue of  $\Sigma$  is  $\Omega(1)$ , so  $v_i = \Sigma^{-1/2}(a_i \otimes a_i)$  has  $\psi_1 \leq O(1)$  and satisfies the boundedness condition (7.3) with  $K \leq O(1)$ .

Thus, we can apply Lemma 7.3 (with  $N = \text{rank } \Sigma \leq n^2$  and  $\varepsilon := c \max(n/\sqrt{m}, n^2/m)$ ) and find that in each case w.h.p.

$$A_{2,2} = \frac{1}{m} \sum_{i=1}^m a_i a_i^T \otimes a_i a_i^T \leq (1 + \varepsilon) \Sigma \leq (1 + \varepsilon) (I + F + \Phi\Phi^T) \quad (7.9)$$

Since  $h_{\mathcal{X}}(Y) \geq 0$  whenever  $Y \geq 0$ , we have  $h_{\mathcal{X}}(A_{2,2}) \leq (1 + \varepsilon)h_{\mathcal{X}}(\Sigma)$ . Additionally,  $h_{\mathcal{X}}(I + F + \Phi\Phi^T) \leq h_{\mathcal{X}}(I) + h_{\mathcal{X}}(F) + h_{\mathcal{X}}(\Phi\Phi^T)$ , so we can bound each of three terms separately. Observe that  $I$  and  $F$  each have largest eigenvalue equal to 1, and so  $h_{\mathcal{X}}(I) \leq 1$  and  $h_{\mathcal{X}}(F) \leq 1$ . (In fact, these are both equalities.)

However, the single nonzero eigenvalue of  $\Phi\Phi^T$  is equal to  $n$ . Here we will need to use the symmetry constraint on  $\mathcal{X}$ . Let  $X^\Gamma$  be the matrix with entries  $X_{(i_1, i_2), (i_3, i_4)}^\Gamma := X_{(i_1, i_4), (i_3, i_2)}$ . If  $X \in \mathcal{X}$  then  $X = X^\Gamma$ . Additionally,  $\langle X, Y \rangle = \langle X^\Gamma, Y^\Gamma \rangle$ . Thus

$$h_{\mathcal{X}}(\Phi\Phi^T) = h_{\mathcal{X}}((\Phi\Phi^T)^\Gamma) \leq \|(\Phi\Phi^T)^\Gamma\|_{2 \rightarrow 2} = 1.$$

This last equality follows from the fact that  $(\Phi\Phi^T)^\Gamma = F$ .

Putting together these ingredients, we obtain the proof of the theorem.  $\square$

It may seem surprising that the factor of  $3^{1/4}$  emerges even for matrices with, say,  $\pm 1$  entries. An intuitive justification for this is that even if the columns of  $A$  are not Gaussian vectors, most linear combinations of them resemble Gaussians. The following Lemma shows that this behavior begins as soon as  $n$  is  $\omega(1)$ .

**Lemma 7.4.** Let  $A = \sum_{i=1}^m e_i a_i^T / \sqrt{n}$  with  $\mathbb{E}_i \|a_i\|_2^4 \geq 1$ . Then  $\|A\|_{2 \rightarrow 4} \geq (3/(1 + 2/n))^{1/4}$ .

To see that the denominator cannot be improved in general, observe that when  $n = 1$  a random sign matrix will have  $2 \rightarrow 4$  norm equal to 1.

*Proof.* Choose  $x \in \mathbb{R}^n$  to be a random Gaussian vector such that  $\mathbb{E}_x \|x\|_2^2 = 1$ . Then

$$\mathbb{E}_x \|Ax\|_4^4 = \mathbb{E}_i \mathbb{E}_x n^{-2} (a_i^T x)^4 = n^2 \mathbb{E}_i \mathbb{E}_x \langle a_i, x \rangle^4 = 3 \mathbb{E}_i \|a_i\|_2^4 \geq 3. \quad (7.10)$$

The last equality comes from the fact that  $\langle a_i, x \rangle$  is a Gaussian random variable with mean zero and variance  $\|a_i\|_2^2/n$ . On the other hand,  $\mathbb{E}_x \|x\|_2^4 = 1 + 2/n$ . Thus, there must exist an  $x$  for which  $\|Ax\|_4^4 / \|x\|_2^4 \geq 3/(1 + 2/n)$ .  $\square$

**Remark 7.5.** It is instructive to consider a variant of the above argument. A simpler upper bound on the value of  $\text{Tensor-SDP}(A)$  is given simply by  $\|A_{2,2}\|$ . However, the presence of the  $\Phi\Phi^T$  term means that this bound will be off by an  $n$ -dependent factor. Thus we observe that the symmetry constraints of  $\text{Tensor-SDP}^{(4)}$  provide a crucial advantage over the simpler bound using eigenvalues. In the language of quantum information (see Section 9.3), this means that the PPT constraint is necessary for the approximation to succeed. See Section 9.3.2 for an example of this that applies to higher levels of the hierarchy as well.

On the other hand, when the  $a_i$  are chosen to be random *complex* Gaussian vectors, we simply have  $\mathbb{E} a_i a_i^* \otimes a_i a_i^* = I + F$ . In this case, the upper bound  $\text{Tensor-SDP}(A) \leq \|A_{2,2}\|$  is already sufficient. Thus, only real random vectors demonstrate a separation between these two bounds.

## 8 The 2-to- $q$ norm and small-set expansion

In this section we show that a graph is a *small-set expander* if and only if the projector to the subspace of its adjacency matrix's top eigenvalues has a bounded  $2 \rightarrow q$  norm for even  $q \geq 4$ . While the “if” part was known before, the “only if” part is novel. This characterization of small-set expanders is of general interest, and also leads to a reduction from the SMALL-SET EXPANSION problem considered in [RS10] to the problem of obtaining a good approximation for the  $2 \rightarrow q$  norms.

**Notation.** For a regular graph  $G = (V, E)$  and a subset  $S \subseteq V$ , we define the *measure* of  $S$  to be  $\mu(S) = |S|/|V|$  and we define  $G(S)$  to be the distribution obtained by picking a random  $x \in S$  and then outputting a random neighbor  $y$  of  $x$ . We define the *expansion* of  $S$ , to be  $\Phi_G(S) = \mathbb{P}_{y \in G(S)}[y \notin S]$ , where  $y$  is a random neighbor of  $x$ . For  $\delta \in (0, 1)$ , we define  $\Phi_G(\delta) = \min_{S \subseteq V: \mu(S) \leq \delta} \Phi_G(S)$ . We often drop the subscript  $G$  from  $\Phi_G$  when it is clear from context. We identify  $G$  with its normalized adjacency (i.e., random walk) matrix. For every  $\lambda \in [-1, 1]$ , we denote by  $V_{\geq \lambda}(G)$  the subspace spanned by the eigenvectors of  $G$  with eigenvalue at least  $\lambda$ . The projector into this subspace is denoted  $P_{\geq \lambda}(G)$ . For a distribution  $D$ , we let  $\text{cp}(D)$  denote the collision probability of  $D$  (the probability that two independent samples from  $D$  are identical).

Our main theorem of this section is the following:

**Theorem** (Restatement of Theorem 2.4). *For every regular graph  $G$ ,  $\lambda > 0$  and even  $q$ ,*

1. (Norm bound implies expansion) *For all  $\delta > 0, \varepsilon > 0$ ,  $\|P_{\geq \lambda}(G)\|_{2 \rightarrow q} \leq \varepsilon / \delta^{(q-2)/2q}$  implies that  $\Phi_G(\delta) \geq 1 - \lambda - \varepsilon^2$ .*

2. (Expansion implies norm bound) *There is a constant  $c$  such that for all  $\delta > 0$ ,  $\Phi_G(\delta) > 1 - \lambda 2^{-cq}$  implies  $\|P_{\geq \lambda}(G)\|_{2 \rightarrow q} \leq 2/\sqrt{\delta}$ .*

One corollary of Theorem 2.4 is that a good approximation to the  $2 \rightarrow 4$  norm implies an approximation of  $\Phi_\delta(G)$ <sup>12</sup>.

**Corollary 8.1.** *If there is a polynomial-time computable relaxation  $\mathcal{R}$  yielding good approximation for the  $2 \rightarrow q$ , then the Small-Set Expansion Hypothesis of [RS10] is false.*

*Proof.* Using [RST10a], to refute the small-set expansion hypothesis it is enough to come up with an efficient algorithm that given an input graph  $G$  and sufficiently small  $\delta > 0$ , can distinguish between the *Yes* case:  $\Phi_G(\delta) < 0.1$  and the *No* case  $\Phi_G(\delta') > 1 - 2^{-c \log(1/\delta')}$  for any  $\delta' \geq \delta$  and some constant  $c$ . In particular for all  $\eta > 0$  and constant  $d$ , if  $\delta$  is small enough then in the *No* case  $\Phi_G(\delta^{0.4}) > 1 - \eta$ . Using Theorem 2.4, in the *Yes* case we know  $\|V_{1/2}(G)\|_{2 \rightarrow 4} \geq 1/(10\delta^{1/4})$ , while in the *No* case, if we choose  $\eta$  to be smaller then  $\eta(1/2)$  in the Theorem, then we know that  $\|V_{1/2}(G)\|_{2 \rightarrow 4} \leq 2/\sqrt{\delta^{0.2}}$ . Clearly, if we have a good approximation for the  $2 \rightarrow 4$  norm then, for sufficiently small  $\delta$  we can distinguish between these two cases.  $\square$

The first part of Theorem 2.4 follows from previous work (e.g., see [KV05]). For completeness, we include a proof in Appendix B. The second part will follow from the following lemma:

**Lemma 8.2.** *Set  $e = e(\lambda, q) := 2^{cq}/\lambda$ , with a constant  $c \leq 100$ . Then for every  $\lambda > 0$  and  $1 \geq \delta \geq 0$ , if  $G$  is a graph that satisfies  $\text{cp}(G(S)) \leq 1/(e|S|)$  for all  $S$  with  $\mu(S) \leq \delta$ , then  $\|f\|_q \leq 2\|f\|_2/\sqrt{\delta}$  for all  $f \in V_{\geq \lambda}(G)$ .*

**Proving the second part of Theorem 2.4 from Lemma 8.2.** We use the variant of the local Cheeger bound obtained in [Ste10, Theorem 2.1], stating that if  $\Phi_G(\delta) \geq 1 - \eta$  then for every  $f \in L_2(V)$  satisfying  $\|f\|_1^2 \leq \delta\|f\|_2^2$ ,  $\|Gf\|_2^2 \leq c\sqrt{\eta}\|f\|_2^2$ . The proof follows by noting that for every set  $S$ , if  $f$  is the characteristic function of  $S$  then  $\|f\|_1 = \|f\|_2^2 = \mu(S)$ , and  $\text{cp}(G(S)) = \|Gf\|_2^2/(\mu(S)|S|)$ .  $\square$

*Proof of Lemma 8.2.* Fix  $\lambda > 0$ . We assume that the graph satisfies the condition of the Lemma with  $e = 2^{cq}/\lambda$ , for a constant  $c$  that we'll set later. Let  $G = (V, E)$  be such a graph, and  $f$  be function in  $V_{\geq \lambda}(G)$  with  $\|f\|_2 = 1$  that maximizes  $\|f\|_q$ . We write  $f = \sum_{i=1}^m \alpha_i \chi_i$  where  $\chi_1, \dots, \chi_m$  denote the eigenfunctions of  $G$  with values  $\lambda_1, \dots, \lambda_m$  that are at least  $\lambda$ . Assume towards a contradiction that  $\|f\|_q > 2/\sqrt{\delta}$ . We'll prove that  $g = \sum_{i=1}^m (\alpha_i/\lambda_i)\chi_i$  satisfies  $\|g\|_q \geq 10\|f\|_q/\lambda$ . This is a contradiction since (using  $\lambda_i \in [\lambda, 1]$ )  $\|g\|_2 \leq \|f\|_2/\lambda$ , and we assumed  $f$  is a function in  $V_{\geq \lambda}(G)$  with a maximal ratio of  $\|f\|_q/\|f\|_2$ .

Let  $U \subseteq V$  be the set of vertices such that  $|f(x)| \geq 1/\sqrt{\delta}$  for all  $x \in U$ . Using Markov and the fact that  $\mathbb{E}_{x \in V}[f(x)^2] = 1$ , we know that  $\mu(U) = |U|/|V| \leq \delta$ , meaning that under our assumptions any subset  $S \subseteq U$  satisfies  $\text{cp}(G(S)) \leq 1/(e|S|)$ . On the other hand, because  $\|f\|_q^q \geq 2^q/\delta^{q/2}$ , we know that  $U$  contributes at least half of the term  $\|f\|_q^q = \mathbb{E}_{x \in V} f(x)^q$ . That is, if we define  $\alpha$  to be  $\mu(U) \mathbb{E}_{x \in U} f(x)^q$  then  $\alpha \geq \|f\|_q^q/2$ . We'll prove the lemma by showing that  $\|g\|_q^q \geq 10\alpha/\lambda$ .

Let  $c$  be a sufficiently large constant ( $c = 100$  will do). We define  $U_i$  to be the set  $\{x \in U : f(x) \in [c^i/\sqrt{\delta}, c^{i+1}/\sqrt{\delta}]\}$ , and let  $I$  be the maximal  $i$  such that  $U_i$  is non-empty.

<sup>12</sup>Note that although we use the  $2 \rightarrow 4$  norm for simplicity, a similar result holds for the  $2 \rightarrow q$  norm for every constant even  $q$ .

Thus, the sets  $U_0, \dots, U_I$  form a partition of  $U$  (where some of these sets may be empty). We let  $\alpha_i$  be the contribution of  $U_i$  to  $\alpha$ . That is,  $\alpha_i = \mu_i \mathbb{E}_{x \in U_i} f(x)^q$ , where  $\mu_i = \mu(U_i)$ . Note that  $\alpha = \alpha_0 + \dots + \alpha_I$ . We'll show that there are some indices  $i_1, \dots, i_J$  such that:

- (i)  $\alpha_{i_1} + \dots + \alpha_{i_J} \geq \alpha / (2c^{10})$ .
- (ii) For all  $j \in [J]$ , there is a non-negative function  $g_j : V \rightarrow \mathbb{R}$  such that  $\mathbb{E}_{x \in V} g_j(x)^q \geq e\alpha_{i_j} / (10c^2)^{q/2}$ .
- (iii) For every  $x \in V$ ,  $g_1(x) + \dots + g_J(x) \leq |g(x)|$ .

Showing these will complete the proof, since it is easy to see that for two non-negative functions and even  $q, g', g''$ ,  $\mathbb{E}(g'(x) + g''(x))^q \geq \mathbb{E} g'(x)^q + \mathbb{E} g''(x)^q$ , and hence (ii) and (iii) imply that

$$\|g\|_4^4 = \mathbb{E} g(x)^4 \geq (e/(10c^2)^{q/2}) \sum_j \alpha_{i_j}. \quad (8.1)$$

Using (i) we conclude that for  $e \geq (10c)^q / \lambda$ , the right-hand side of (8.1) will be larger than  $10\alpha/\lambda$ .

We find the indices  $i_1, \dots, i_J$  iteratively. We let  $\mathcal{I}$  be initially the set  $\{0..I\}$  of all indices. For  $j = 1, 2, \dots$  we do the following as long as  $\mathcal{I}$  is not empty:

1. Let  $i_j$  be the largest index in  $\mathcal{I}$ .
2. Remove from  $\mathcal{I}$  every index  $i$  such that  $\alpha_i \leq c^{10}\alpha_{i_j}/2^{i-i_j}$ .

We let  $J$  denote the step when we stop. Note that our indices  $i_1, \dots, i_J$  are sorted in descending order. For every step  $j$ , the total of the  $\alpha_i$ 's for all indices we removed is less than  $c^{10}\alpha_{i_j}$  and hence we satisfy (i). The crux of our argument will be to show (ii) and (iii). They will follow from the following claim:

**Claim 8.3.** *Let  $S \subseteq V$  and  $\beta > 0$  be such that  $|S| \leq \delta$  and  $|f(x)| \geq \beta$  for all  $x \in S$ . Then there is a set  $T$  of size at least  $e|S|$  such that  $\mathbb{E}_{x \in T} g(x)^2 \geq \beta^2/4$ .*

The claim will follow from the following lemma:

**Lemma 8.4.** *Let  $D$  be a distribution with  $\text{cp}(D) \leq 1/N$  and  $g$  be some function. Then there is a set  $T$  of size  $N$  such that  $\mathbb{E}_{x \in T} g(x)^2 \geq (\mathbb{E} g(D))^2/4$ .*

*Proof.* Identify the support of  $D$  with the set  $[M]$  for some  $M$ , we let  $p_i$  denote the probability that  $D$  outputs  $i$ , and sort the  $p_i$ 's such that  $p_1 \geq p_2 \geq \dots \geq p_M$ . We let  $\beta'$  denote  $\mathbb{E} g(D)$ ; that is,  $\beta' = \sum_{i=1}^M p_i g(i)$ . We separate to two cases. If  $\sum_{i>N} p_i g(i) \geq \beta'/2$ , we define the distribution  $D'$  as follows: we set  $\mathbb{P}[D' = i]$  to be  $p_i$  for  $i > N$ , and we let all  $i \leq N$  be equiprobable (that is be output with probability  $(\sum_{i=1}^N p_i)/N$ ). Clearly,  $\mathbb{E} |g(D')| \geq \sum_{i>N} p_i g(i) \geq \beta'/2$ , but on the other hand, since the maximum probability of any element in  $D'$  is at most  $1/N$ , it can be expressed as a convex combination of flat distributions over sets of size  $N$ , implying that one of these sets  $T$  satisfies  $\mathbb{E}_{x \in T} |g(x)| \geq \beta'/2$ , and hence  $\mathbb{E}_{x \in T} g(x)^2 \geq \beta'^2/4$ .

The other case is that  $\sum_{i=1}^N p_i g(i) \geq \beta'/2$ . In this case we use Cauchy-Schwarz and argue that

$$\beta'^2/4 \leq \left( \sum_{i=1}^N p_i^2 \right) \left( \sum_{i=1}^N g(i)^2 \right). \quad (8.2)$$

But using our bound on the collision probability, the right-hand side of (8.2) is upper bounded by  $\frac{1}{N} \sum_{i=1}^N g(i)^2 = \mathbb{E}_{x \in [N]} g(x)^2$ .  $\square$



*Proof of Claim 8.3 from Lemma 8.4.* By construction  $f = Gg$ , and hence we know that for every  $x$ ,  $f(x) = \mathbb{E}_{y \sim x} g(y)$ . This means that if we let  $D$  be the distribution  $G(S)$  then

$$\mathbb{E} |g(D)| = \mathbb{E}_{x \in S} \mathbb{E}_{y \sim x} |g(y)| \geq \mathbb{E}_{x \in S} |\mathbb{E}_{y \sim x} g(y)| = \mathbb{E}_{x \in S} |f(x)| \geq \beta .$$

By the expansion property of  $G$ ,  $\text{cp}(D) \leq 1/(e|S|)$  and thus by Lemma 8.4 there is a set  $T$  of size  $e|S|$  satisfying  $\mathbb{E}_{x \in T} g(x)^2 \geq \beta^2/4$ .  $\square$

We will construct the functions  $g_1, \dots, g_J$  by applying iteratively Claim 8.3. We do the following for  $j = 1, \dots, J$ :

1. Let  $T_j$  be the set of size  $e|U_{i_j}|$  that is obtained by applying Claim 8.3 to the function  $f$  and the set  $U_{i_j}$ . Note that  $\mathbb{E}_{x \in T_j} g(x)^2 \geq \beta_{i_j}^2/4$ , where we let  $\beta_i = c^i / \sqrt{\delta}$  (and hence for every  $x \in U_i$ ,  $\beta_i \leq |f(x)| \leq c\beta_i$ ).
2. Let  $g'_j$  be the function on input  $x$  that outputs  $\gamma \cdot |g(x)|$  if  $x \in T_j$  and 0 otherwise, where  $\gamma \leq 1$  is a scaling factor that ensures that  $\mathbb{E}_{x \in T_j} g'_j(x)^2$  equals exactly  $\beta_{i_j}^2/4$ .
3. We define  $g_j(x) = \max\{0, g'_j(x) - \sum_{k < j} g_k(x)\}$ .

Note that the second step ensures that  $g'_j(x) \leq |g(x)|$ , while the third step ensures that  $g_1(x) + \dots + g_j(x) \leq g'_j(x)$  for all  $j$ , and in particular  $g_1(x) + \dots + g_J(x) \leq |g(x)|$ . Hence the only thing left to prove is the following:

**Claim 8.5.**  $\mathbb{E}_{x \in V} g_j(x)^q \geq e\alpha_{i_j}/(10c)^{q/2}$

*Proof.* Recall that for every  $i$ ,  $\alpha_i = \mu_i \mathbb{E}_{x \in U_i} f(x)^q$ , and hence (using  $f(x) \in [\beta_i, c\beta_i]$  for  $x \in U_i$ ):

$$\mu_i \beta_i^q \leq \alpha_i \leq \mu_i c^q \beta_i^q . \quad (8.3)$$

Now fix  $T = T_j$ . Since  $\mathbb{E}_{x \in V} g_j(x)^q$  is at least (in fact equal)  $\mu(T) \mathbb{E}_{x \in T} g_j(x)^q$  and  $\mu(T) = e\mu(U_{i_j})$ , we can use (8.3) and  $\mathbb{E}_{x \in T} g_j(x)^q \geq (E_{x \in T} g_j(x)^2)^{q/2}$ , to reduce proving the claim to showing the following:

$$\mathbb{E}_{x \in T} g_j(x)^2 \geq (c\beta_{i_j})^2 / (10c^2) = \beta_{i_j}^2 / 10 . \quad (8.4)$$

We know that  $\mathbb{E}_{x \in T} g'_j(x)^2 = \beta_{i_j}^2/4$ . We claim that (8.4) will follow by showing that for every  $k < j$ ,

$$\mathbb{E}_{x \in T} g'_k(x)^2 \leq 100^{-i'} \cdot \beta_{i_j}^2 / 4 , \quad (8.5)$$

where  $i' = i_k - i_j$ . (Note that  $i' > 0$  since in our construction the indices  $i_1, \dots, i_J$  are sorted in descending order.)

Indeed, (8.5) means that if we let momentarily  $\|g_j\|$  denote  $\sqrt{\mathbb{E}_{x \in T} g_j(x)^2}$  then

$$\|g_j\| \geq \|g'_j\| - \|\sum_{k < j} g_k\| \geq \|g'_j\| - \sum_{k < j} \|g_k\| \geq \|g'_j\| (1 - \sum_{i'=1}^{\infty} 10^{-i'}) \geq 0.8 \|g'_j\| . \quad (8.6)$$

The first inequality holds because we can write  $g_j$  as  $g'_j - h_j$ , where  $h_j = \min\{g'_j, \sum_{k < j} g_k\}$ . Then, on the one hand,  $\|g_j\| \geq \|g'_j\| - \|h_j\|$ , and on the other hand,  $\|h_j\| \leq \|\sum_{k < j} g_k\|$  since  $g'_j \geq 0$ . The second inequality holds because  $\|g_k\| \leq \|g'_k\|$ . By squaring (8.6) and plugging in the value of  $\|g'_j\|^2$  we get (8.4).

**Proof of (8.5).** By our construction, it must hold that

$$c^{10} \alpha_{i_k} / 2^{i'} \leq \alpha_{i_j}, \quad (8.7)$$

since otherwise the index  $i_j$  would have been removed from the  $\mathcal{I}$  at the  $k^{\text{th}}$  step. Since  $\beta_{i_k} = \beta_{i_j} c^{i'}$ , we can plug (8.3) in (8.7) to get

$$\mu_{i_k} c^{10+4i'} / 2^{i'} \leq c^4 \mu_{i_j}$$

or

$$\mu_{i_k} \leq \mu_{i_j} (2/c)^{4i'} c^{-6}.$$

Since  $|T_i| = e|U_i|$  for all  $i$ , it follows that  $|T_k|/|T| \leq (2/c)^{4i'} c^{-6}$ . On the other hand, we know that  $\mathbb{E}_{x \in T_k} g'_k(x)^2 = \beta_{i_k}^2 / 4 = c^{2i'} \beta_{i_j}^2 / 4$ . Thus,

$$\mathbb{E}_{x \in T} g'_k(x)^2 \leq 2^{4i'} c^{2i'-4i'-6} \beta_{i_j}^2 / 4 \leq (2^4/c^2)^{i'} \beta_{i_j}^2 / 4,$$

and now we just choose  $c$  sufficiently large so that  $c^2/2^4 > 100$ . □

□

## 9 Relating the 2-to-4 norm and the injective tensor norm

In this section, we present several equivalent formulations of the 2-to-4 norm: 1) as the injective tensor norm of a 4-tensor, 2) as the injective tensor norm of a 3-tensor, and 3) as the maximum of a linear function over a convex set, albeit a set where the weak membership problem is hard. Additionally, we can consider maximizations over real or complex vectors. These equivalent formulations are discussed in Section 9.1.

We use this to show hardness of approximation (Theorem 2.5) for the 2-to-4 norm in Section 9.2, and then show positive algorithmic results (Theorem 2.3) in Section 9.3. Somewhat surprisingly, many of the key arguments in these sections are imported from the quantum information literature, even though no quantum algorithms are involved. It is an interesting question to find a more elementary proof of the result in Section 9.3.

We will generally work with the counting norms  $\|\cdot\|$ , defined as  $\|x\|_p := (\sum_i |x_i|^p)^{1/p}$ , and the counting inner product, defined by  $\langle x, y \rangle := x^* y$ , where  $*$  denotes the conjugate transpose.

### 9.1 Equivalent maximizations

#### 9.1.1 Injective tensor norm and separable states

Recall from the introduction the definition of the injective tensor norm: if  $V_1, \dots, V_r$  are vector spaces with  $T \in V_1 \otimes \dots \otimes V_r$ , then  $\|T\|_{\text{inj}} = \max\{|\langle T, (x_1 \otimes \dots \otimes x_r) \rangle| : x_1 \in \mathbf{S}(V_1), \dots, x_r \in \mathbf{S}(V_r)\}$ , where  $\mathbf{S}(V)$  denotes the  $L_2$ -unit vectors in a vector space  $V$ . In this paper we use the term “injective tensor norm” to mean the injective tensor norm of  $\ell_2$  spaces, and we caution the reader that in other contexts it has a more general meaning. These norms were introduced by Grothendieck, and they are further discussed in [Rya02].

We will also need the definition of separable states from quantum information. For a vector space  $V$ , define  $L(V)$  to be the linear operators on  $V$ , and define  $\mathcal{D}(V) := \{\rho \in L(V) : \rho \geq 0, \text{Tr } \rho = 1\} = \text{conv}\{uv^* : v \in \mathbf{S}(V)\}$  to be the *density operators* on  $V$ . The trace induces

an inner product on operators:  $\langle X, Y \rangle := \text{Tr } X^* Y$ . An important class of density operators are the *separable density operators*. For vector spaces  $V_1, \dots, V_r$ , these are

$$\text{Sep}(V_1, \dots, V_r) := \text{conv} \left\{ v_1 v_1^* \otimes \dots \otimes v_r v_r^* : \forall i, v_i \in \mathbf{S}(V_i) \right\}.$$

If  $V = V_1 = \dots = V_r$ , then let  $\text{Sep}^r(V)$  denote  $\text{Sep}(V_1, \dots, V_r)$ . Physically, density operators are the quantum analogues of probability distributions, and separable density operators describe unentangled quantum states; conversely, entangled states are defined to be the set of density operators that are not separable. For readers familiar with quantum information, we point out that our treatment differs principally in its use of the expectation for norms and inner products, rather than the sum.

For any bounded convex set  $K$ , define the *support function* of  $K$  to be

$$\mathbf{h}_K(x) := \max_{y \in K} |\langle x, y \rangle|.$$

Define  $e_i \in \mathbb{F}^n$  to be the vector with 1 in the  $i^{\text{th}}$  position. Now we can give the convex-optimization formulation of the injective tensor norm.

**Lemma 9.1.** *Let  $V_1, \dots, V_r$  be vector spaces with  $n_i := \dim V_i$ , and  $T \in V_1 \otimes \dots \otimes V_r$ . Choose an orthonormal basis  $e_1, \dots, e_{n_r}$  for  $V_r$ . Define  $T_1, \dots, T_{n_r} \in V_1 \otimes \dots \otimes V_{r-1}$  by  $T = \sum_{i=1}^{n_r} T_i \otimes e_i$  and define  $M \in L(V_1 \otimes \dots \otimes V_{r-1})$  by  $M = \sum_{i=1}^{n_r} T_i T_i^*$ . Then*

$$\|T\|_{\text{inj}}^2 = \mathbf{h}_{\text{Sep}(V_1, \dots, V_{r-1})}(M). \quad (9.1)$$

Observe that any  $M \geq 0$  can be expressed in this form, possibly by padding  $n_r$  to be at least rank  $M$ . Thus calculating  $\|\cdot\|_{\text{inj}}$  for  $r$ -tensors is equivalent in difficulty to computing  $\mathbf{h}_{\text{Sep}^1(\mathbb{F}^{n_1})}$  for p.s.d. arguments. This argument appeared before in [HM10], where it was explained using quantum information terminology.

It is instructive to consider the  $r = 2$  case. In this case,  $T$  is equivalent to a matrix  $\hat{T}$  and  $\|T\|_{\text{inj}} = \|\hat{T}\|_{2 \rightarrow 2}$ . Moreover  $\text{Sep}^1(\mathbb{F}^{n_1}) = \mathcal{D}(\mathbb{F}^{n_1})$  is simply the convex hull of  $vv^*$  for unit vectors  $v$ . Thus  $\mathbf{h}_{\text{Sep}^1(\mathbb{F}^{n_1})}(M)$  is simply the maximum eigenvalue of  $M = TT^*$ . In this case, Lemma 9.1 merely states that the square of the largest singular value of  $\hat{T}$  is the largest eigenvalue of  $\hat{T}\hat{T}^*$ . The general proof follows this framework.

*Proof of Lemma 9.1.*

$$\|T\|_{\text{inj}} = \max_{x_1 \in \mathbf{S}(V_1), \dots, x_r \in \mathbf{S}(V_r)} |\langle T, x_1 \otimes \dots \otimes x_r \rangle| \quad (9.2)$$

$$= \max_{x_1 \in \mathbf{S}(V_1), \dots, x_{r-1} \in \mathbf{S}(V_{r-1})} \max_{x_r \in \mathbf{S}(V_r)} \left| \sum_{i=1}^{n_r} \langle T_i, x_1 \otimes \dots \otimes x_{r-1} \rangle \cdot \langle e_i, x_r \rangle \right| \quad (9.3)$$

$$= \max_{x_1 \in \mathbf{S}(V_1), \dots, x_{r-1} \in \mathbf{S}(V_{r-1})} \left\| \sum_{i=1}^{n_r} \langle T_i, x_1 \otimes \dots \otimes x_{r-1} \rangle e_i \right\|_2 \quad (9.4)$$

Therefore

$$\|T\|_{\text{inj}}^2 = \max_{x_1 \in \mathbf{S}(V_1), \dots, x_{r-1} \in \mathbf{S}(V_{r-1})} \left\| \sum_{i=1}^{n_r} \langle T_i, x_1 \otimes \dots \otimes x_{r-1} \rangle e_i \right\|_2^2 \quad (9.5)$$

$$= \max_{x_1 \in \mathbf{S}(V_1), \dots, x_{r-1} \in \mathbf{S}(V_{r-1})} \sum_{i=1}^{n_r} |\langle T_i, x_1 \otimes \dots \otimes x_{r-1} \rangle|_2^2 \quad (9.6)$$

$$= \max_{x_1 \in \mathbf{S}(V_1), \dots, x_{r-1} \in \mathbf{S}(V_{r-1})} \left\langle \sum_{i=1}^{n_r} T_i T_i^*, x_1 x_1^* \otimes \dots \otimes x_r x_r^* \right\rangle \quad (9.7)$$

$$= \mathbf{h}_{\text{Sep}(V_1, \dots, V_{r_1})} \left( \sum_{i=1}^{n_r} T_i T_i^* \right) \quad (9.8)$$

□

In what follows, we will also need to make use of some properties of symmetric tensors. Define  $\mathcal{S}_k$  to be the group of permutations of  $[k]$  and define  $P_n(\pi) \in L((\mathbb{F}^n)^{\otimes k})$  to be the operator that permutes  $k$  tensor copies of  $\mathbb{F}^n$  according to  $\pi$ . Formally,

$$P_n(\pi) := \sum_{i_1, \dots, i_r \in [d]} \bigotimes_{k=1}^r e_{i_k} e_{i_{\pi(k)}}^T. \quad (9.9)$$

Then define  $\vee^k \mathbb{F}^n$  to be the subspace of vectors in  $(\mathbb{F}^n)^{\otimes r}$  that are unchanged by each  $P_n(\pi)$ . This space is called the *symmetric subspace*. A classic result in symmetric polynomials states that  $\vee^r \mathbb{F}^n$  is spanned by the vectors  $\{v^{\otimes r} : v \in \mathbb{F}^n\}$ .<sup>13</sup>

One important fact about symmetric tensors is that for injective tensor norm, the vectors in the maximization can be taken to be equal. Formally,

**Fact 9.2.** *If  $T \in \vee^r \mathbb{F}^n$  then*

$$\|T\|_{\text{inj}} = \max_{x \in \mathbf{S}(\mathbb{F}^n)} |\langle T, x^{\otimes r} \rangle|. \quad (9.10)$$

This has been proven in several different works; see the paragraph above Eq. (3.1) of [CKP00] for references.

### 9.1.2 Connection to the 2-to-4 norm

Let  $A = \sum_{i=1}^m e_i a_i^T$ , so that  $a_1, \dots, a_m \in \mathbb{R}^n$  are the rows of  $A$ . Define

$$A_4 = \sum_{i=1}^m a_i^{\otimes 4} \in (\mathbb{R}^n)^{\otimes 4} \quad (9.11)$$

$$A_3 = \sum_{i=1}^m a_i \otimes a_i \otimes e_i \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^m \quad (9.12)$$

$$A_{2,2} = \sum_{i=1}^m a_i a_i^T \otimes a_i a_i^T \in L((\mathbb{R}^n)^{\otimes 2}) \quad (9.13)$$

The subscripts indicate that that  $A_r$  is an  $r$ -tensor, and  $A_{r,s}$  is a map from  $r$ -tensors to  $s$ -tensors.

Further, for a real tensor  $T \in (\mathbb{R}^n)^{\otimes r}$ , define  $\|T\|_{\text{inj}[\mathbb{C}]}$  to be the injective tensor norm that results from treating  $T$  as a complex tensor; that is,  $\max\{|\langle T, x_1 \otimes \dots \otimes x_r \rangle| : x_1, \dots, x_r \in \mathbf{S}(\mathbb{C}^n)\}$ . For  $r \geq 3$ ,  $\|T\|_{\text{inj}[\mathbb{C}]}$  can be larger than  $\|T\|_{\text{inj}}$  by as much as  $\sqrt{2}$  [CKP00].

Our main result on equivalent forms of the 2  $\rightarrow$  4 norm is the following.

**Lemma 9.3.**

$$\|A\|_{2 \rightarrow 4}^4 = \|A_4\|_{\text{inj}} = \|A_3\|_{\text{inj}}^2 = \|A_4\|_{\text{inj}[\mathbb{C}]} = \|A_3\|_{\text{inj}[\mathbb{C}]}^2 = \mathbf{h}_{\text{Sep}^2(\mathbb{R}^n)}(A_{2,2}) = \mathbf{h}_{\text{Sep}^2(\mathbb{C}^n)}(A_{2,2})$$

<sup>13</sup>For the proof, observe that  $v^{\otimes r} \in \vee^r \mathbb{F}^n$  for any  $v \in \mathbb{F}^n$ . To construct a basis for  $\vee^r \mathbb{F}^n$  out of linear combinations of different  $v^{\otimes r}$ , let  $z_1, \dots, z_n$  be indeterminates and evaluate the  $r$ -fold derivatives of  $(z_1 e_1 + \dots + z_n e_n)^{\otimes r}$  at  $z_1 = \dots = z_n = 0$ .

*Proof.*

$$\|A\|_{2 \rightarrow 4}^4 = \max_{x \in \mathbf{S}(\mathbb{R}^n)} \sum_{i=1}^m \langle a_i, x \rangle^4 \quad (9.14)$$

$$= \max_{x \in \mathbf{S}(\mathbb{R}^n)} \langle A_4, x^{\otimes 4} \rangle \quad (9.15)$$

$$= \max_{x_1, x_2, x_3, x_4 \in \mathbf{S}(\mathbb{R}^n)} |\langle A_4, x_1 \otimes x_2 \otimes x_3 \otimes x_4 \rangle| \quad (9.16)$$

$$= \|A_4\|_{\text{inj}} \quad (9.17)$$

Here (9.16) follows from Fact 9.2.

Next one can verify with direct calculation (and using  $\max_{z \in \mathbf{S}(\mathbb{R}^n)} \langle v, z \rangle = \|v\|_2$ ) that

$$\max_{x \in \mathbf{S}(\mathbb{R}^n)} \langle A_4, x^{\otimes 4} \rangle = \max_{x \in \mathbf{S}(\mathbb{R}^n)} \langle A_{2,2}, xx^T \otimes xx^T \rangle = \max_{x \in \mathbf{S}(\mathbb{R}^n)} \max_{z \in \mathbf{S}(\mathbb{R}^m)} \langle A_3, x \otimes x \otimes z \rangle^2. \quad (9.18)$$

Now define  $z(i) := \langle e_i, z \rangle$  and continue.

$$\max_{x \in \mathbf{S}(\mathbb{R}^n)} \max_{z \in \mathbf{S}(\mathbb{R}^m)} |\langle A_3, x \otimes x \otimes z \rangle| = \max_{x \in \mathbf{S}(\mathbb{R}^n)} \max_{z \in \mathbf{S}(\mathbb{R}^m)} \operatorname{Re} \sum_{i=1}^m z(i) \langle a_i, x \rangle^2 \quad (9.19)$$

$$= \max_{x \in \mathbf{S}(\mathbb{R}^n)} \max_{z \in \mathbf{S}(\mathbb{C}^m)} \operatorname{Re} \sum_{i=1}^m z(i) \langle a_i, x \rangle^2 \quad (9.20)$$

$$= \max_{z \in \mathbf{S}(\mathbb{C}^m)} \left\| \sum_{i=1}^m z(i) a_i a_i^T \right\|_{2 \rightarrow 2} \quad (9.21)$$

$$= \max_{z \in \mathbf{S}(\mathbb{C}^m)} \max_{x, y \in \mathbf{S}(\mathbb{C}^n)} \operatorname{Re} \sum_{i=1}^m z(i) \langle x^*, a_i \rangle \langle a_i, y \rangle \quad (9.22)$$

$$= \|A_3\|_{\text{inj}[\mathbb{C}]} = \|A_3\|_{\text{inj}} \quad (9.23)$$

From Lemma 9.1, we thus have  $\|A\|_{2 \rightarrow 4}^4 = \mathbf{h}_{\text{Sep}^2(\mathbb{R}^n)}(A_{2,2}) = \mathbf{h}_{\text{Sep}^2(\mathbb{C}^n)}(A_{2,2})$ .

To justify (9.22), we argue that the maximum in (9.21) is achieved by taking all the  $z(i)$  real (and indeed nonnegative). The resulting matrix  $\sum_i z(i) a_i a_i^T$  is real and symmetric, so its operator norm is achieved by taking  $x = y$  to be real vectors. Thus, the maximum in  $\|A_3\|_{\text{inj}[\mathbb{C}]}$  is achieved for real  $x, y, z$  and as a result  $\|A_3\|_{\text{inj}[\mathbb{C}]} = \|A_3\|_{\text{inj}}$ .

Having now made the bridge to complex vectors, we can work backwards to establish the last equivalence:  $\|A_4\|_{\text{inj}[\mathbb{C}]}$ . Repeating the argument that led to (9.17) will establish that  $\|A_4\|_{\text{inj}[\mathbb{C}]} = \max_{x \in \mathbf{S}(\mathbb{C}^n)} \max_{z \in \mathbf{S}(\mathbb{C}^m)} |\langle A_3, x \otimes x \otimes z \rangle|^2 = \|A_3\|_{\text{inj}[\mathbb{C}]}^2$ .  $\square$

## 9.2 Hardness of approximation for the 2-to-4 norm

This section is devoted to the proof of Theorem 2.5, establishing hardness of approximation for the 2-to-4 norm.

First, we restate Theorem 2.5 more precisely. We omit the reduction to when  $A$  is a projector, deferring this argument to Corollary 9.9, where we will further use a randomized reduction.

**Theorem 9.4.** (restatement of Theorem 2.5) *Let  $\phi$  be a 3-SAT instance with  $n$  variables and  $O(n)$  clauses. Determining whether  $\phi$  is satisfiable can be reduced in polynomial time to determining whether  $\|A\|_{2 \rightarrow 4} \geq C$  or  $\|A\|_{2 \rightarrow 4} \leq c$  where  $0 \leq c < C$  and  $A$  is an  $m \times m$  matrix. This is possible for two choices of parameters:*

1.  $m = \text{poly}(n)$ , and  $C/c > 1 + 1/n \text{poly} \log(n)$ ; or,
2.  $m = \exp(\sqrt{n} \text{poly} \log(n) \log(C/c))$ .

The key challenge is establishing the following reduction.

**Lemma 9.5.** *Let  $M \in L(\mathbb{C}^n \otimes \mathbb{C}^n)$  satisfy  $0 \leq M \leq I$ . Assume that either (case Y)  $\mathbf{h}_{\text{Sep}(n,n)}(M) = 1$  or (case N)  $\mathbf{h}_{\text{Sep}(n,n)}(M) \leq 1 - \delta$ . Let  $k$  be a positive integer. Then there exists a matrix  $A$  of size  $n^{4k} \times n^{2k}$  such that in case Y,  $\|A\|_{2 \rightarrow 4} = 1$ , and in case N,  $\|A\|_{2 \rightarrow 4} = (1 - \delta/2)^k$ . Moreover,  $A$  can be constructed efficiently from  $M$ .*

*Proof of Theorem 9.4.* Once Lemma 9.5 is proved, Theorem 2.5 follows from previously known results about the hardness of approximating  $\mathbf{h}_{\text{Sep}}$ . Let  $\phi$  be a 3-SAT instance with  $n$  variables and  $O(n)$  clauses. In Theorem 4 of [GNN] (improving on earlier work of [Gur03]), it was proved that  $\phi$  can be reduced to determining whether  $\mathbf{h}_{\text{Sep}(n^c, n^c)}(M)$  is equal to 1 (“case Y”) or  $\leq 1 - 1/n \log^c(n)$  (“case N”), where  $c > 0$  is a universal constant, and  $M$  is an efficiently constructible matrix with  $0 \leq M \leq I$ . Now we apply Lemma 9.5 with  $k = 1$  to find that exists a matrix  $A$  of dimension  $\text{poly}(n)$  such that in case Y,  $\|A\|_{2 \rightarrow 4} = 1$ , and in case N,  $\|A\|_{2 \rightarrow 4} \leq 1 - 1/2n \log^c(n)$ . Thus, distinguishing these cases would determine whether  $\phi$  is satisfiable. This establishes part (1) of Theorem 2.5.

For part (2), we start with Corollary 14 of [HM10], which gives a reduction from determining the satisfiability of  $\phi$  to distinguishing between (“case Y”)  $\mathbf{h}_{\text{Sep}(m,m)}(M) = 1$  and (“case N”)  $\mathbf{h}_{\text{Sep}(m,m)}(M) \leq 1/2$ . Again  $0 \leq M \leq I$ , and  $M$  can be constructed in time  $\text{poly}(m)$  from  $\phi$ , but this time  $m = \exp(\sqrt{n} \text{poly} \log(n))$ . Applying Lemma 9.5 in a similar fashion completes the proof.  $\square$

*Proof of Lemma 9.5.* The previous section shows that computing  $\|A\|_{2 \rightarrow 4}$  is equivalent to computing  $\mathbf{h}_{\text{Sep}(n,n)}(A_{2,2})$ , for  $A_{2,2}$  defined as in (9.13). However, the hardness results of [Gur03, GNN, HM10] produce matrices  $M$  that are not in the form of  $A_{2,2}$ . The reduction of [HM10] comes closest, by producing a matrix that is a sum of terms of the form  $xx^* \otimes yy^*$ . However, we need a sum of terms of the form  $xx^* \otimes xx^*$ . This will be achieved by a variant of the protocol used in [HM10].

Let  $M_0 \in L(\mathbb{C}^n \otimes \mathbb{C}^n)$  satisfy  $0 \leq M_0 \leq I$ . Consider the promise problem of distinguishing the cases  $\mathbf{h}_{\text{Sep}(n,n)}(M_0) = 1$  (called “case Y”) from  $\mathbf{h}_{\text{Sep}(n,n)}(M_0) \leq 1/2$  (called “case N”). We show that this reduces to finding a multiplicative approximation for  $\|A\|_{2 \rightarrow 4}$  for some real  $A$  of dimension  $n^\alpha$  for a constant  $\alpha > 0$ . Combined with known hardness-of-approximation results (Corollary 15 of [HM10]), this will imply Theorem 2.5.

Define  $P$  to be the projector onto the subspace of  $(\mathbb{C}^n)^{\otimes 4}$  that is invariant under  $P_n((1, 3))$  and  $P_n((2, 4))$  (see Section 9.1 for definitions). This can be obtained by applying  $P_n((2, 3))$  to  $\vee^2 \mathbb{C}^n \otimes \vee^2 \mathbb{C}^n$ , where we recall that  $\vee^2 \mathbb{C}^n$  is the symmetric subspace of  $(\mathbb{C}^n)^{\otimes 2}$ . Since  $P$  projects onto the vectors invariant under the 4-element group generated by  $P_n((1, 3))$  and  $P_n((2, 4))$ , we can write it as

$$P = \frac{I + P_n((1, 3))}{2} \cdot \frac{I + P_n((2, 4))}{2}. \quad (9.24)$$

An alternate definition of  $P$  is due to Wick’s theorem:

$$P = \mathbb{E}_{a,b} [aa^* \otimes bb^* \hat{\otimes} aa^* \otimes bb^*], \quad (9.25)$$

where the expectation is taken over complex-Gaussian-distributed vectors  $a, b \in \mathbb{C}^n$  normalized so that  $\mathbb{E} \|a\|_2^2 = \mathbb{E} \|b\|_2^2 = n/\sqrt{2}$ . Here we use the notation  $\hat{\otimes}$  to mark the separation

between systems that we will use to define the separable states  $\text{Sep}(n^2, n^2)$ . We could equivalently write  $P = \mathbb{E}_{a,b}[(aa^* \otimes bb^*)^{\otimes 2}]$ . We will find that (9.24) is more useful for doing calculations, while (9.25) is helpful for converting  $M_0$  into a form that resembles  $A_{2,2}$  for some matrix  $A$ .

Define  $M_1 = (\sqrt{M_0} \hat{\otimes} \sqrt{M_0})P(\sqrt{M_0} \hat{\otimes} \sqrt{M_0})$ , where  $\sqrt{M_0}$  is taken to be the unique positive-semidefinite square root of  $M_0$ . Observe that

$$M_1 = \mathbb{E}_{a,b}[v_{a,b}v_{a,b}^* \hat{\otimes} v_{a,b}v_{a,b}^*] = \mathbb{E}_{a,b}[V_{a,b}^{\otimes 2}], \quad (9.26)$$

where we define  $v_{a,b} := \sqrt{M_0}(a \otimes b)$  and  $V_{a,b} := v_{a,b}v_{a,b}^*$ . We claim that  $\mathbf{h}_{\text{Sep}}(M_1)$  gives a reasonable proxy for  $\mathbf{h}_{\text{Sep}}(M_0)$  in the following sense.

**Lemma 9.6.**

$$\mathbf{h}_{\text{Sep}(n^2, n^2)}(M_1) \begin{cases} = 1 & \text{in case Y} \\ \leq 1 - \delta/2 & \text{in case N.} \end{cases} \quad (9.27)$$

The proof of Lemma 9.6 is deferred to the end of this section. The analysis is very similar to Theorem 13 of [HM10], but the analysis here is much simpler because  $M_0$  acts on only two systems. However, it is strictly speaking not a consequence of the results in [HM10], because that paper considered a slightly different choice of  $M_1$ .

The advantage of replacing  $M_0$  with  $M_1$  is that (thanks to (9.25)) we now have a matrix with the same form as  $A_{2,2}$  in (9.13), allowing us to make use of Lemma 9.3. However, we first need to amplify the separation between cases Y and N. This is achieved by the matrix  $M_2 := M_1^{\otimes k}$ . This tensor product is *not* across the cut we use to define separable states; in other words:

$$M_2 = \mathbb{E}_{\substack{a_1, \dots, a_k \\ b_1, \dots, b_k}} [(V_{a_1, b_1} \otimes \dots \otimes V_{a_k, b_k})^{\otimes 2}]. \quad (9.28)$$

Now Lemma 12 from [HM10] implies that  $\mathbf{h}_{\text{Sep}(n^{2k}, n^{2k})}(M_2) = \mathbf{h}_{\text{Sep}(n^2, n^2)}(M_1)^k$ . This is either 1 or  $\leq (3/4)^k$ , depending on whether we have case Y or N.

Finally, we would like to relate this to the  $2 \rightarrow 4$  norm of a matrix. It will be more convenient to work with  $M_1$ , and then take tensor powers of the corresponding matrix. Naively applying Lemma 9.3 would relate  $\mathbf{h}_{\text{Sep}}(M_1)$  to  $\|A\|_{2 \rightarrow 4}$  for an infinite-dimensional  $A$ . Instead, we first replace the continuous distribution on  $a$  (resp.  $b$ ) with a finitely-supported distribution in a way that does not change  $\mathbb{E}_a aa^* \otimes aa^*$  (resp.  $\mathbb{E}_b bb^* \otimes bb^*$ ). Such distributions are called complex-projective (2,2)-designs or quantum (state) 2-designs, and can be constructed from spherical 4-designs on  $\mathbb{R}^{2n}$  [AE07]. Finding these designs is challenging when each vector needs to have the same weight, but for our purposes we can use Carathéodory's theorem to show that there exist vectors  $z_1, \dots, z_m$  with  $m = n^2$  such that

$$\mathbb{E}_a [aa^* \otimes aa^*] = \sum_{i \in [m]} z_i z_i^* \otimes z_i z_i^*. \quad (9.29)$$

In what follows, assume that the average over  $a, b$  used in the definitions of  $P, M_1, M_2$  is replaced by the sum over  $z_1, \dots, z_m$ . By (9.29) this change does not affect the values of  $P, M_1, M_2$ .

For  $i, j \in [m]$ , define  $w_{i,j} := \sqrt{M_0}(z_i \otimes z_j)$ , and let  $e_{i,j} := e_i \otimes e_j$ . Now we can apply Lemma 9.3 to find that  $\mathbf{h}_{\text{Sep}}(M_1) = \|A_1\|_{2 \rightarrow 4}^4$ , where

$$A_1 = \sum_{i,j \in [m]} e_{i,j} w_{i,j}^*.$$

The amplified matrix  $M_2$  similarly satisfies  $\mathbf{h}_{\text{Sep}(n^{2k}, n^{2k})}(M_2) = \|A_2\|_{2 \rightarrow 4}^4$ , where

$$A_2 := A_1^{\otimes k} = \sum_{i_1, \dots, i_k, j_1, \dots, j_k \in [m]} (e_{i_1, j_1} \otimes \dots \otimes e_{i_k, j_k})(w_{i_1, j_1} \otimes \dots \otimes w_{i_k, j_k})^*.$$

The last step is to relate the complex matrix  $A_2$  to a real matrix  $A_3$  with the same  $2 \rightarrow 4$  norm once we restrict to real inputs. This can be achieved by replacing a single complex entry  $\alpha + i\beta$  with the  $6 \times 2$  real matrix

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 2^{1/4} & 0 \\ 2^{1/4} & 0 \\ 0 & 2^{1/4} \\ 0 & 2^{1/4} \end{pmatrix} \cdot \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$$

A complex input  $x + iy$  is represented by the column vector  $\begin{pmatrix} x \\ y \end{pmatrix}$ . The initial  $2 \times 2$  matrix maps this to the real representation of  $(\alpha + i\beta)(x + iy)$ , and then the fixed  $6 \times 2$  matrix maps this to a vector whose 4-norm equals  $|(\alpha + i\beta)(x + iy)|^4$ .  $\square$

We conclude with the proof of Lemma 9.6, mostly following [HM10].

*Proof.* Case Y is simplest, and also provides intuition for the choices of the  $M_1$  construction. Since the extreme points of  $\text{Sep}(n, n)$  are of the form  $xx^* \otimes yy^*$  for  $x, y \in \mathbf{S}(\mathbb{C}^n)$ , it follows that there exists  $x, y \in \mathbf{S}(\mathbb{C}^n)$  with  $\langle x \otimes y, M(x \otimes y) \rangle = 1$ . Since  $M \leq I$ , this implies that  $M(x \otimes y) = (x \otimes y)$ . Thus  $\sqrt{M_0}(x \otimes y) = (x \otimes y)$ . Let

$$z = x \otimes y \otimes x \otimes y.$$

Then  $z$  is an eigenvector of both  $\sqrt{M_0} \otimes \sqrt{M_0}$  and  $P$ , with eigenvalue 1 in each case. To see this for  $P$ , we use the definition in (9.24). Thus  $\langle z, M_1 z \rangle = 1$ , and it follows that  $\mathbf{h}_{\text{Sep}(n^2, n^2)}(M_1) \geq 1$ . On the other hand,  $M_1 \leq I$ , implying that  $\mathbf{h}_{\text{Sep}(n^2, n^2)}(M_1) \leq 1$ . This establishes case Y.

For case N, we assume that  $\mathbf{h}_{\text{Sep}(n, n)}(M_0) \leq 1 - \delta$  for any  $x, y \in \mathbf{S}(\mathbb{C}^n)$ . The idea of the proof is that for any  $x, y \in \mathbf{S}(\mathbb{C}^{n^2})$ , we must either have  $x, y$  close to a product state, in which case the  $\sqrt{M_0}$  step will shrink the vector, or if they are far from a product state and preserved by  $\sqrt{M_0} \otimes \sqrt{M_0}$ , then the  $P$  step will shrink the vector. In either case, the length will be reduced by a dimension-independent factor.

We now spell this argument out in detail. Choose  $x, y \in \mathbf{S}(\mathbb{C}^{n^2})$  to achieve

$$s := \langle x \otimes y, M_1(x \otimes y) \rangle = \mathbf{h}_{\text{Sep}(n^2, n^2)}(M_1). \quad (9.30)$$

Let  $X, Y \in L(\mathbb{C}^n)$  be defined by

$$\sqrt{M_0}x =: \sum_{i, j \in [n]} X_{i, j} e_i \otimes e_j \quad \text{and} \quad \sqrt{M_0}y =: \sum_{i, j \in [n]} Y_{i, j} e_i \otimes e_j \quad (9.31)$$

Note that  $\langle X, X \rangle = \langle x, M_0 x \rangle \leq 1$  and similarly for  $\langle Y, Y \rangle$ . We wish to estimate

$$s = \sum_{i, j, k, l, i', j', k', l' \in [n]} \bar{X}_{i', j'} \bar{Y}_{k', l'} X_{i, j} Y_{k, l} \langle e_{i'} \otimes e_{j'} \otimes e_{k'} \otimes e_{l'}, P(e_i \otimes e_j \otimes e_k \otimes e_l) \rangle \quad (9.32)$$



Using (9.24) we see that the expression inside the  $\langle \cdot \rangle$  is

$$\frac{\delta_{i,i'}\delta_{j,j'}\delta_{k,k'}\delta_{l,l'} + \delta_{i,k'}\delta_{j,j'}\delta_{k,i'}\delta_{l,l'} + \delta_{i,i'}\delta_{j,l'}\delta_{k,k'}\delta_{l,j'} + \delta_{i,k'}\delta_{j,l'}\delta_{k,i'}\delta_{l,j'}}{4} \quad (9.33)$$

Rearranging, we find

$$s = \frac{\langle X, X \rangle \langle Y, Y \rangle + \langle X, Y \rangle \langle X, Y \rangle + \langle YY^*, XX^* \rangle + \langle Y^*Y, X^*X \rangle}{4}. \quad (9.34)$$

Using the AM-GM inequality we see that the maximum of this expression is achieved when  $X = Y$ , in which case we have

$$s = \frac{\langle X, X \rangle^2 + \langle X^*X, X^*X \rangle}{2} \leq \frac{1 + \langle X^*X, X^*X \rangle}{2}. \quad (9.35)$$

Let the singular values of  $X$  be  $\sigma_1 \geq \dots \geq \sigma_n$ . Observe that  $\|\sigma\|_2^2 = \langle X, X \rangle \leq 1$ , and thus  $\|\sigma\|_4^4 = \langle X^*X, X^*X \rangle \leq \sigma_1^2$ . On the other hand,

$$\sigma_1^2 = \max_{a,b \in \mathbb{S}(\mathbb{C}^n)} |\langle a, Xb \rangle|^2 \quad (9.36)$$

$$= \max_{a,b \in \mathbb{S}(\mathbb{C}^n)} |\langle a \otimes b, \sqrt{M_0}x \rangle|^2 \quad (9.37)$$

$$= \max_{a,b \in \mathbb{S}(\mathbb{C}^n)} |\langle \sqrt{M_0}(a \otimes b), x \rangle|^2 \quad (9.38)$$

$$= \max_{a,b \in \mathbb{S}(\mathbb{C}^n)} \langle \sqrt{M_0}(a \otimes b), \sqrt{M_0}(a \otimes b) \rangle \quad (9.39)$$

$$= \max_{a,b \in \mathbb{S}(\mathbb{C}^n)} \langle a \otimes b, M_0(a \otimes b) \rangle \quad (9.40)$$

$$= \mathbf{h}_{\text{Sep}(n,n)}(M_0) \leq 1 - \delta \quad (9.41)$$

□

**Remark:** It is possible to extend Lemma 9.5 to the situation when case  $Y$  has  $\mathbf{h}_{\text{Sep}}(M) > 1 - \delta'$  for some constant  $\delta' < \delta$ . Since the details are somewhat tedious, and repeat arguments in [HM10], we omit them here.

## 9.2.1 Hardness of approximation for projectors

Can Theorem 2.5 give any super-polynomial lower bound for the SSE problem if we assume the Exponential-Time Hypothesis for 3-SAT? To resolve this question using our techniques, we would like to reduce 3-SAT to estimating the  $2 \rightarrow 4$  norm of the projector onto the eigenvectors of a graph that have large eigenvalue. We do not know how to do this. However, instead, we show that the matrix  $A$  constructed in Theorem 2.5 can be taken to be a projector. This is almost WLOG, except that the resulting  $2 \rightarrow 4$  norm will be at least  $3^{1/4}$ .

**Lemma 9.7.** *Let  $A$  be a linear map from  $\mathbb{R}^k$  to  $\mathbb{R}^n$  and  $0 < c < C$ ,  $\varepsilon > 0$  some numbers. Then there is  $m = O(n^2/\varepsilon^2)$  and a map  $A'$  from  $\mathbb{R}^k$  to  $\mathbb{R}^m$  such that  $\sigma_{\min}(A') \geq 1 - \varepsilon$  and (i) if  $\|A\|_{2 \rightarrow 4} \leq c$  then  $\|A'\|_{2 \rightarrow 4} \leq 3^{1/4} + \varepsilon$ , (ii)  $\|A\|_{2 \rightarrow 4} \geq C$  then  $\|A'\|_{2 \rightarrow 4} \geq \Omega(\varepsilon C/c)$ .*

*Proof.* We let  $B$  be a random map from  $\mathbb{R}^k$  to  $\mathbb{R}^{O(n^2/\delta^2)}$  with entries that are i.i.d. Gaussians with mean zero and variance  $1/\sqrt{k}$ . Then Dvoretzky's theorem [Pis99] implies that for every  $f \in \mathbb{R}^k$ ,  $\|Bf\|_4 \in 3^{1/4}(1 \pm \delta)\|f\|_2$ . Consider the operator  $A' = \begin{pmatrix} A \\ B \end{pmatrix}$  that maps  $f$  into the

concatenation of  $Af$  and  $Bf$ . Moreover we take multiple copies of each coordinate so that the measure of output coordinates of  $A'$  corresponding to  $A$  is  $\alpha = \delta/c^4$ , while the measure of coordinates corresponding to  $B$  is  $1 - \alpha$ .

Now for every function  $f$ , we get that  $\|A'f\|_4^4 = \alpha\|Af\|_4^4 + (1 - \alpha)\|Bf\|_4^4$ . In particular, since  $\|Bf\|_4^4 \in 3(1 \pm \delta)\|f\|_2^4$ , we get that if  $f$  is a unit vector and  $\|Af\|_4^4 \leq c^4$  then  $\|A'f\|_4^4 \leq \delta^{1/4} + 3(1 + \delta)$ , while if  $\|Af\|_4^4 \geq C^4$ , then  $\|A'f\|_4^4 \geq \delta(C/c)^4$ .

Also note that the random operator  $B$  will satisfy that for every function  $f$ ,  $\|Bf\|_2 \geq (1 - \delta)\|f\|_2$ , and hence  $\|A'f\| \geq (1 - \alpha)(1 - \delta)\|f\|$ . Choosing  $\delta = \varepsilon/2$  concludes the proof.  $\square$

It turns out that for the purposes of hardness of good approximation, the case that  $A$  is a projector is almost without loss of generality.

**Lemma 9.8.** *Suppose that for some  $\varepsilon > 0, C > 1 + \varepsilon$  there is a  $\text{poly}(n)$  algorithm that on input a subspace  $V \subseteq \mathbb{R}^n$  can distinguish between the case (Y)  $\|\Pi_V\|_{2 \rightarrow 4} \geq C$  and the case (N)  $\|\Pi_V\|_{2 \rightarrow 4} \leq 3^{1/4} + \varepsilon$ , where  $\Pi_V$  denotes the projector onto  $V$ . Then there is  $\delta = \Omega(\varepsilon)$  and a  $\text{poly}(n)$  algorithm that on input an operator  $A : \mathbb{R}^k \rightarrow \mathbb{R}^n$  with  $\sigma_{\min}(A) \geq 1 - \delta$  can distinguish between the case (Y)  $\|A\|_{2 \rightarrow 4} \geq C(1 + \delta)$  and (N)  $\|A\|_{2 \rightarrow 4} \leq 3^{1/4}(1 + \delta)$ .*

*Proof.* First we can assume without loss of generality that  $\|A\|_{2 \rightarrow 2} = \sigma_{\max}(A) \leq 1 + \delta$ , since otherwise we could rule out case (N). Now we let  $V$  be the image of  $A$ . In the case (N) we get that that for every  $f \in \mathbb{R}^k$

$$\|Af\|_4 \leq 3^{1/4}(1 + \delta)\|f\|_2 \leq 3^{1/4}(1 + \delta)\|Af\|_2 / \sigma_{\min}(A) \leq 3^{1/4}(1 + O(\delta))\|Af\|_2,$$

implying  $\|\Pi_V\|_{2 \rightarrow 4} \leq 3^{1/4} + O(\delta)$ . In the case (Y) we get that there is some  $f$  such that  $\|Af\|_4 \geq C(1 + \delta)\|f\|_2$ , but since  $\|Af\|_2 \leq \sigma_{\max}(A)\|f\|_2$ , we get that  $\|Af\|_4 \geq C$ , implying  $\|\Pi_V\|_{2 \rightarrow 4} \geq C$ .  $\square$

Together these two lemmas effectively extend Theorem 2.5 to the case when  $A$  is a projector. We focus on the hardness of approximating to within a constant factor.

**Corollary 9.9.** *For any  $\ell, \varepsilon > 0$ , if  $\phi$  is a 3-SAT instance with  $n$  variables and  $O(n)$  clauses, then determining satisfiability of  $\phi$  can be reduced to distinguishing between the cases  $\|A\|_{2 \rightarrow 4} \leq 3^{1/4} + \varepsilon$  and  $\|A\|_{2 \rightarrow 4} \geq \ell$ , where  $A$  is a projector acting on  $m = \exp(\sqrt{n} \text{poly} \log(n) \log(\ell/\varepsilon))$  dimensions.*

*Proof.* Start as in the proof of Theorem 2.5, but in the application of Lemma 9.5, take  $k = O(\log(\ell/\varepsilon))$ . This will allow us to take  $C/c = \Omega(\ell/\varepsilon)$  in Lemma 9.7. Translating into a projector with Lemma 9.8, we obtain the desired result.  $\square$

### 9.3 Algorithmic applications of equivalent formulations

In this section we discuss the positive algorithmic results that come from the equivalences in Section 9.1. Since entanglement plays such a central role in quantum mechanics, the set  $\text{Sep}^2(\mathbb{C}^n)$  has been extensively studied. However, because its hardness has long been informally recognized (and more recently has been explicitly established [Gur03, Liu07, HM10, GNN]), various relaxations have been proposed for the set. These relaxations are generally efficiently computable, but also have limited accuracy; see [BS10] for a review.

Two of the most important relaxations are the PPT condition and  $k$ -extendability. For an operator  $X \in L((\mathbb{C}^n)^{\otimes r})$  and a set  $S \subseteq [r]$ , define the *partial transpose*  $X^{T_S}$  to be the result of applying the transpose map to the systems  $S$ . Formally, we define

$$(X_1 \otimes \cdots \otimes X_r)^{T_S} := \bigotimes_{k=1}^r f_k(X_k)$$

$$f_k(M) := \begin{cases} M & \text{if } k \notin S \\ M^T & \text{if } k \in S \end{cases}$$

and extend  $T_S$  linearly to all of  $L((\mathbb{C}^n)^{\otimes r})$ . One can verify that if  $X \in \text{Sep}^r(\mathbb{C}^n)$  then  $X^{T_S} \geq 0$  for all  $S \subseteq [r]$ . In this case we say that  $X$  is PPT, meaning that it has Positive Partial Transposes. However, the converse is not always true. If  $n > 2$  or  $r > 2$ , then there are states which are PPT but not in  $\text{Sep}$  [HHH96].

The second important relaxation of  $\text{Sep}$  is called  $r$ -extendability. To define this, we need to introduce the partial trace. For  $S \subseteq [r]$ , we define  $\text{Tr}_S$  to be the map from  $L((\mathbb{C}^n)^{\otimes r})$  to  $L((\mathbb{C}^n)^{\otimes r-|S|})$  that results from applying  $\text{Tr}$  to the systems in  $S$ . Formally

$$\text{Tr}_S \bigotimes_{k=1}^r X_k = \prod_{k \in S} \text{Tr} X_k \bigotimes_{k \notin S} X_k,$$

and  $\text{Tr}_S$  extends by linearity to all of  $L((\mathbb{C}^n)^{\otimes r})$ .

To obtain our relaxation of  $\text{Sep}$ , we say that  $\rho \in \mathcal{D}(\mathbb{C}^n \otimes \mathbb{C}^n)$  is  $r$ -extendable if there exists a symmetric extension  $\sigma \in \mathcal{D}(\mathbb{C}^n \otimes \vee^r \mathbb{C}^n)$  such that  $\text{Tr}_{\{3, \dots, r+1\}} \sigma = \rho$ . Observe that if  $\rho \in \text{Sep}^2(\mathbb{C}^n)$ , then we can write  $\rho = \sum_i x_i x_i^* \otimes y_i y_i^*$ , and so  $\sigma = \sum_i x_i x_i^* \otimes (y_i y_i^*)^{\otimes r}$  is a valid symmetric extension. Thus the set of  $k$ -extendable states contains the set of separable states, but again the inclusion is strict. Indeed, increasing  $k$  gives an infinite hierarchy of strictly tighter approximations of  $\text{Sep}^2(\mathbb{C}^n)$ . This hierarchy ultimately converges [DPS04], although not always at a useful rate (see Example IV.1 of [CKMR07]). Interestingly this relaxation is known to completely fail as a method of approximating  $\text{Sep}^2(\mathbb{R}^n)$  [CFS02], but our Lemma 9.3 is evidence that those difficulties do not arise in the  $2 \rightarrow 4$ -norm problem.

These two relaxations can be combined to optimize over symmetric extensions that have positive partial transposes [DPS04]. Call this the *level- $r$  DPS relaxation*. It is known to converge in some cases more rapidly than  $r$ -extendability alone [NOP09], but also is never exact for any finite  $r$  [DPS04]. Like SoS, this relaxation is an SDP with size  $n^{O(r)}$ . In fact, for the case of the  $2 \rightarrow 4$  norm, the relaxations are equivalent.

**Lemma 9.10.** *When the level- $r$  DPS relaxation is applied to  $A_{2,2}$ , the resulting approximation is equivalent to Tensor-SDP $^{(2r+2)}$*

*Proof.* Suppose we are given an optimal solution to the level- $r$  DPS relaxation. This can be thought of as a density operator  $\sigma \in \mathcal{D}(\mathbb{C}^n \otimes \vee^r \mathbb{C}^n)$  whose objective value is  $\lambda := \langle A_{2,2}, \text{Tr}_{\{3, \dots, r+1\}} \sigma \rangle = \langle A_{2,2} \otimes I_n^{\otimes r-1}, \sigma \rangle$ . Let  $\Pi_{\text{sym}}^{(2)} := (I + P_n((1, 2)))/2$  be the orthogonal projector onto  $\vee^2 \mathbb{C}^n$ . Then  $A_{2,2} = \Pi_{\text{sym}}^{(2)} A_{2,2} \Pi_{\text{sym}}^{(2)}$ . Thus, we can replace  $\sigma$  by  $\sigma' := (\Pi_{\text{sym}}^{(2)} \otimes I_n^{\otimes r-1}) \sigma (\Pi_{\text{sym}}^{(2)} \otimes I_n^{\otimes r-1})$  without changing the objective function. However, unless  $\sigma' = \sigma$ , we will have  $\text{Tr} \sigma' < 1$ . In this case, either  $\sigma' = 0$  and  $\lambda = 0$ , or  $\sigma' / \text{Tr} \sigma'$  is a solution of the DPS relaxation with a higher objective value. In either case, this contradicts the assumption that  $\lambda$  is the optimal value. Thus, we must have  $\sigma = \sigma'$ , and in particular  $\text{supp} \sigma \subseteq \vee^2 \mathbb{C}^n \otimes (\mathbb{C}^n)^{\otimes r-1}$ . Since we had  $\text{supp} \sigma \subseteq \mathbb{C}^n \otimes \vee^r \mathbb{C}^n$  by assumption, it follows that

$$\text{supp} \sigma \subseteq (\vee^2 \mathbb{C}^n \otimes (\mathbb{C}^n)^{\otimes r-1}) \cap (\mathbb{C}^n \otimes \vee^r \mathbb{C}^n) = \vee^{r+1} \mathbb{C}^n$$

Observe next that  $\sigma^T$  is also a valid and optimal solution to the DPS relaxation, and so  $\sigma' = (\sigma + \sigma^T)/2$  is as well. Since  $\sigma'$  is both symmetric and Hermitian, it must be a real matrix. Replacing  $\sigma$  with  $\sigma'$ , we see that we can assume WLOG that  $\sigma$  is real.

Similarly, the PPT condition implies that  $\sigma^{T_A} \geq 0$ . (Recall that the first system is  $A$  and the rest are  $B_1, \dots, B_k$ .) Since the partial transpose doesn't change the objective

function,  $\sigma' = (\sigma + \sigma^{T_A})/2$  is also an optimal solution. Replacing  $\sigma$  with  $\sigma'$ , we see that we can assume WLOG that  $\sigma = \sigma^{T_A}$ . Let  $\vec{\sigma} \in (\mathbb{R}^n)^{\otimes 2r+2}$  denote the flattening of  $\sigma$ ; i.e.  $\langle x \otimes y, \vec{\sigma} \rangle = \langle x, \sigma y \rangle$  for all  $x, y \in (\mathbb{R}^n)^{r+1}$ . Then the fact that  $\sigma = \sigma^{T_A}$  means that  $\vec{\sigma}$  is invariant under the action of  $P_n((1, r+1))$ . Similarly, the fact that  $\text{supp } \sigma \subseteq \vee^{r+1} \mathbb{R}^n$  implies that  $\vec{\sigma} \in \vee^{r+1} \mathbb{R}^n \otimes \vee^{r+1} \mathbb{R}^n$ . Combining these two facts we find that  $\vec{\sigma} \in \vee^{2r+2} \mathbb{R}^n$ .

Now that  $\vec{\sigma}$  is fully symmetric under exchange of all  $2r+2$  indices, we can interpret it as a real-valued pseudo-expectation  $\tilde{\mathbb{E}}_\sigma$  for polynomials of degree  $2r+2$ . More precisely, we can define the linear map *coeff* that sends homogeneous degree- $2r+2$  polynomials to  $\vee^{2r+2} \mathbb{R}^n$  by its action on monomials:

$$\text{coeff}(f_1^{\alpha_1} \cdots f_n^{\alpha_n}) := \Pi_{\text{sym}}^{(2r+2)}(e_1^{\otimes \alpha_1} \otimes \cdots \otimes e_n^{\otimes \alpha_n}), \quad (9.42)$$

where  $\Pi_{\text{sym}}^{(2r+2)} := \frac{1}{2r+2!} \sum_{\pi \in \mathcal{S}_{2r+2}} P_n(\pi)$ . For a homogenous polynomial  $Q(f)$  of even degree  $2r' \leq 2r+2$  we define *coeff* by

$$\text{coeff}(Q(f)) := \text{coeff}(Q(f) \cdot \|f\|_2^{2r+2-2r'}).$$

For a homogenous polynomial  $Q(f)$  of odd degree, we set  $\text{coeff}(Q) := 0$ . Then we can extend *coeff* by linearity to all polynomials of degree  $\leq 2r+2$ . Now define

$$\tilde{\mathbb{E}}_\sigma[Q] := \langle \text{coeff}(Q), \vec{\sigma} \rangle.$$

We claim that this is a valid pseudo-expectation. For normalization, observe that  $\tilde{\mathbb{E}}[1] = \langle \text{coeff}(\|f\|_2^{2r+2}), \vec{\sigma} \rangle = \text{Tr } \sigma = 1$ . Similarly, the Tensor-SDP constraint of  $\tilde{\mathbb{E}}[(\|f\|_2^2 - 1)^2] = 0$  is satisfied by our definition of *coeff*. Linearity follows from the linearity of *coeff* and the inner product. For positivity, consider a polynomial  $Q(f)$  of degree  $\leq r+1$ . Write  $Q = Q_o + Q_e$ , where  $Q_o$  collects all monomials of odd degree and  $Q_e$  collects all monomials of even degree (i.e.  $Q_e, Q_o = (Q(f) \pm Q(-f))/2$ ). Then  $\tilde{\mathbb{E}}[Q^2] = \tilde{\mathbb{E}}[Q_o^2] + \tilde{\mathbb{E}}[Q_e^2]$ , using the property that the pseudo-expectation of a monomial of odd degree is zero.

Consider first  $\tilde{\mathbb{E}}[Q_e^2]$ . Let  $r' = 2\lfloor \frac{r+1}{2} \rfloor$  (i.e.  $r'$  is  $r+1$  rounded down to the nearest even number), so that  $Q_e = \sum_{i=0}^{r'/2} Q_{2i}$ , where  $Q_{2i}$  is homogenous of degree  $2i$ . Define  $Q'_e := \sum_{i=0}^{r'/2} Q_{2i} \|f\|_2^{r'-2i}$ . Observe that  $Q'_e$  is homogenous of degree  $r' \leq r+1$ , and that  $\tilde{\mathbb{E}}[Q_e^2] = \tilde{\mathbb{E}}[(Q'_e)^2]$ . Next, define *coeff'* to map homogenous polynomials of degree  $r'$  into  $\vee^{r'} \mathbb{R}^n$  by replacing  $2r+2$  in (9.42) with  $r'$ . If  $r' = r+1$  then define  $\sigma' = \sigma$ , or if  $r' = r$  then define  $\sigma' = \text{Tr}_A \sigma$ . Thus  $\sigma'$  acts on  $r'$  systems. Define  $\vec{\sigma}' \in \vee^{2r'} \mathbb{R}^n$  to be the flattened version of  $\sigma'$ . Finally we can calculate

$$\tilde{\mathbb{E}}[Q_e^2] = \tilde{\mathbb{E}}[(Q'_e)^2] = \langle \text{coeff}'(Q'_e) \otimes \text{coeff}'(Q'_e), \vec{\sigma}' \rangle = \langle \text{coeff}' Q'_e, \sigma' \text{coeff}' Q'_e \rangle \geq 0.$$

A similar argument establishes that  $\tilde{\mathbb{E}}[Q_o^2] \geq 0$  as well. This establishes that any optimal solution to the DPS relaxation translates into a solution of the Tensor-SDP relaxation.

To translate a Tensor-SDP solution into a DPS solution, we run this construction in reverse. The arguments are essentially the same, except that we no longer need to establish symmetry across all  $2r+2$  indices.  $\square$

### 9.3.1 Approximation guarantees and the proof of Theorem 2.3

Many approximation guarantees for the  $k$ -extendable relaxation (with or without the additional PPT constraints) required that  $k$  be poly( $n$ ), and thus do not lead to useful algorithms. Recently, [BaCY11] showed that in some cases it sufficed to take  $k = O(\log n)$ , leading

to quasi-polynomial algorithms. It is far from obvious that their proof translates into our sum-of-squares framework, but nevertheless Lemma 9.10 implies that Tensor-SDP can take advantage of their analysis.

To apply the algorithm of [BaCY11], we need to upper-bound  $A_{2,2}$  by an 1-LOCC measurement operator. That is, a quantum measurement that can be implemented by one-way Local Operations and Classical Communication (LOCC). Such a measurement should have a decomposition of the form  $\sum_i V_i \otimes W_i$  where each  $V_i, W_i \geq 0$ ,  $\sum_i V_i \leq I_n$  and each  $W_i \leq I_n$ . Thus, for complex vectors  $v_1, \dots, v_m, w_1, \dots, w_m$  satisfying  $\sum_i v_i v_i^* \leq I_n$  and  $\forall i, w_i w_i^* \leq I_n$ , the operator  $\sum_i v_i v_i^* \otimes w_i w_i^*$  is a 1-LOCC measurement.

To upper-bound  $A_{2,2}$  by a 1-LOCC measurement, we note that  $a_i a_i^T \leq \|a_i\|_2^2 I_n$ . Thus, if we define  $Z := \|\sum_i a_i a_i^T\|_{2 \rightarrow 2} \max_i \|a_i\|_2^2$ , then  $A_{2,2}/Z$  is a 1-LOCC measurement. Note that this is a stricter requirement than merely requiring  $A_{2,2}/Z \leq I_{n^2}$ . On the other hand, in some cases (e.g.  $a_i$  all orthogonal), it may be too pessimistic.

In terms of the original matrix  $A = \sum_i e_i e_i^T$ , we have  $\max_i \|a_i\|_2 = \|A\|_{2 \rightarrow \infty}$ . Also  $\|\sum_i a_i a_i^T\|_{2 \rightarrow 2} = \|A^T A\|_{2 \rightarrow 2} = \|A\|_{2 \rightarrow 2}^2$ . Thus

$$Z = \|A\|_{2 \rightarrow 2}^2 \|A\|_{2 \rightarrow \infty}^2.$$

Recall from the introduction that  $Z$  is an upper bound on  $\|A\|_{2 \rightarrow 4}^4$ , based on the fact that  $\|x\|_4 \leq \sqrt{\|x\|_2 \|x\|_\infty}$  for any  $x$ . (This bound also arises from using interpolation of norms [Ste56].)

We can now apply the argument of [BaCY11] and show that optimizing over  $O(r)$ -extendable states will approximate  $\|A\|_{2 \rightarrow 4}^4$  up to additive error  $\sqrt{\frac{\log(n)}{r}} Z$ . Equivalently, we can obtain additive error  $\varepsilon Z$  using  $O(\log(n)/\varepsilon^2)$ -round Tensor-SDP. Whether the relaxation used is the DPS relaxation or our SoS-based Tensor-SDP algorithm, the resulting runtime is  $\exp(O(\log^2(n)/\varepsilon^2))$ .

### 9.3.2 Gap instances

Since Tensor-SDP is equivalent than the DPS relaxation for separable states, any gap instance for Tensor-SDP would translate into a gap instance for the DPS relaxation. This would mean the existence of a state that passes the  $k$ -extendability and PPT test, but nevertheless is far from separable, with  $A_{2,2}$  serving as the entanglement witness demonstrating this. While such states are already known [DPS04, BS10], it would be of interest to find new such families of states, possibly with different scaling of  $r$  and  $n$ .

Our results, though, can be used to give an asymptotic separation of the DPS hierarchy from the  $r$ -extendability hierarchy. (As a reminder, the DPS hierarchy demands that a state not only have an extension to  $r+1$  parties, but also that the extension be PPT across any cut.) To state this more precisely, we introduce some notation. Define  $\text{DPS}_r$  to be the set of states  $\rho^{AB}$  for which there exists an extension  $\tilde{\rho}^{AB_1 \dots B_r}$  with support in  $\mathbb{C}^n \otimes \vee^r \mathbb{C}^n$  (i.e. a symmetric extension) such that  $\tilde{\rho}$  is invariant under taking the partial transpose of any system. Define  $\text{Ext}_r$  to be the set of states on  $AB$  with symmetric extensions to  $AB_1 \dots B_r$ , but without any requirement about the partial transpose. Both  $\mathbf{h}_{\text{DPS}_r}$  and  $\mathbf{h}_{\text{Ext}_r}$  can be computed in time  $n^{O(r)}$ , although in practice  $\mathbf{h}_{\text{Ext}_r}(M)$  is easier to work with, since it only requires computing the top eigenvalue of  $M \otimes I_n^{\otimes r-1}$  restricted to  $\mathbb{C}^n \otimes \vee^r \mathbb{C}^n$  and does not require solving an SDP.

Many of the results about the convergence of  $\text{DPS}_r$  to Sep (such as [DPS04, CKMR07, KM09, BaCY11]) use only the fact that  $\text{DPS}_r \subset \text{Ext}_r$ . A rare exception is [NOP09], which shows that  $\text{DPS}_r$  is at least quadratically closer to Sep than  $\text{Ext}_r$  is, in the regime where  $r \gg n$ . Another simple example comes from  $M = \Phi \Phi^*$ , where  $\Phi$  is the maximally entangled

state  $n^{-1/2} \sum_{i=1}^n e_i \otimes e_i$ . Then one can readily compute that  $\mathbf{h}_{\text{Sep}}(M) = \mathbf{h}_{\text{DPS}_1}(M) = 1/n$ , while the  $r$ -extendible state

$$\tilde{\rho}^{AB_1 \dots B_r} = \frac{1}{r} \sum_{i=1}^r (\Phi \Phi^*)^{AB_i} \otimes \bigotimes_{j \in [r] \setminus \{i\}} \left( \frac{I}{n} \right)^{B_j} \quad (9.43)$$

achieves  $\mathbf{h}_{\text{Ext}_r}(M) \geq 1/r$ . (In words, (9.43) describes a state where  $A$  and a randomly chosen  $B_i$  share the state  $\Phi \Phi^*$ , while the other  $B_j$  systems are described by maximally mixed states.) This proves that the  $r$ -extendible hierarchy cannot achieve a good multiplicative approximation of  $\mathbf{h}_{\text{Sep}}(M)$  for all  $M$  without taking  $r \geq \Omega(n)$ .

Can we improve this when  $M$  is in a restricted class, such as 1-LOCC? Here [BRSdW11] show that the Khot-Vishnoi integrality construction can yield an  $n^2$ -dimensional  $M$  for which  $h_{\text{Sep}}(M) \leq O(1/n)$ , but  $\text{Tr } M \Phi \geq \Omega(1/\log^2(n))$ . Combined with (9.43) this implies that  $\mathbf{h}_{\text{Ext}_r}(M) \geq \Omega(1/r \log^2(n))$ . On the other hand, Theorem 6.12 and Lemma 9.10 implies that  $\mathbf{h}_{\text{DPS}_3}(M) \leq O(1/n)$ . Additionally, the  $M$  from Ref. [BRSdW11] belongs to the class BELL, a subset of 1-LOCC, given by measurements of the form  $\sum_{i,j} p_{i,j} A_i \otimes B_j$ , with  $0 \leq p_{i,j} \leq 1$  and  $\sum_i A_i = \sum_j B_j = I$ . As a result, we obtain the following corollary.

**Corollary 9.11.** *There exists an  $n^2$  dimensional  $M \in \text{BELL}$  such that*

$$\frac{h_{\text{Ext}_r}(M)}{h_{\text{DPS}_3}(M)} \leq O\left(\frac{r \log^2(n)}{n}\right)$$

## 10 Subexponential algorithm for the 2-to-q norm

In this section we prove Theorem 2.1:

**Theorem** (Restatement of Theorem 2.1). *For every  $1 < c < C$ , there is a  $\text{poly}(n) \exp(n^{2/q})$ -time algorithm that computes a  $(c, C)$ -approximation for the  $2 \rightarrow q$  norm of any linear operator whose range is  $\mathbb{R}^n$ .*

and obtain as a corollary a subexponential algorithm for SMALL-SET EXPANSION. The algorithm roughly matches the performance of [ABS10]'s for the same problem, and in fact is a very close variant of it. The proof is obtained by simply noticing that a subspace  $V$  cannot have too large of a dimension without containing a vector  $v$  (that can be easily found) such that  $\|v\|_q \gg \|v\|_2$ , while of course it is always possible to find such a vector (if it exists) in time exponential in  $\dim(V)$ . The key observation is the following basic fact (whose proof we include here for completeness):

**Lemma 10.1.** *For every subspace  $V \subseteq \mathbb{R}^n$ ,  $\|V\|_{2 \rightarrow \infty} \geq \sqrt{\dim(V)}$ .*

*Proof.* Let  $f^1, \dots, f^d$  be an orthonormal basis for  $V$ , where  $d = \dim(V)$ . For every  $i \in [n]$ , let  $g^i$  be the function  $\sum_{j=1}^d f_i^j f^j$ . Note that the  $i^{\text{th}}$  coordinate of  $g^i$  is equal to  $\sum_{j=1}^d (f_i^j)^2$  (\*) which also equals  $\|g^i\|_2^2$  since the  $f^j$ 's are an orthonormal basis. Also the expectation of (\*) over  $i$  is  $\sum_{j=1}^d \mathbb{E}_{i \in [n]} (f_i^j)^2 = \sum_{j=1}^d \|f^j\|_2^2 = d$  since these are unit vectors. Thus we get that  $\mathbb{E}_i \|g^i\|_\infty \geq \mathbb{E}_i g_i^i = d = \mathbb{E}_i \|g^i\|_2^2$ . We claim that one of the  $g^i$ 's must satisfy  $\|g^i\|_\infty \geq \sqrt{d} \|g^i\|_2$ . Indeed, suppose otherwise, then we'd get that

$$d = \mathbb{E}_i \|g^i\|_2^2 > \mathbb{E}_i \|g^i\|_\infty^2 / d$$

meaning  $\mathbb{E}_i \|g^i\|_\infty^2 < d^2$ , but  $\mathbb{E}_i \|g^i\|_\infty^2 \geq (\mathbb{E}_i \|g^i\|_\infty)^2 = d^2$ —a contradiction.  $\square$

**Corollary 10.2.** For every subspace  $V \subseteq \mathbb{R}^n$ ,  $\|V\|_{2 \rightarrow q} \geq \sqrt{\dim(V)}/n^{1/q}$

*Proof.* By looking at the contribution to the  $q^{\text{th}}$ -norm of just one coordinate one can see that for every function  $f$ ,  $\|f\|_q \geq (\|f\|_\infty^q/n)^{1/q} = \|f\|_\infty/n^{1/q}$ .  $\square$

**Proof of Theorem 2.1 from Corollary 10.2.** Let  $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$  be an operator, and let  $1 < c < C$  be some constants and  $\sigma = \sigma_{\min}(A)$  be such that  $\|Af\|_2 \geq \sigma\|f\|_2$  for every  $f$  orthogonal to the kernel of  $A$ . We want to distinguish between the case that  $\|A\|_{2 \rightarrow q} \leq c$  and the case that  $\|A\|_{2 \rightarrow q} \geq C$ . If  $\sigma > c$  then clearly we are not in the first case, and so we are done. Let  $V$  be the image of  $A$ . If  $\dim(V) \leq C^2 n^{2/q}$  then we can use brute force enumeration to find out if such  $v$  exists in the space. Otherwise, by Corollary 10.2 we must be in the second case.  $\square$

Note that by applying Theorem 2.3 we can replace the brute force enumeration step by the SoS hierarchy, since  $\|V\|_{2 \rightarrow 2} \leq 1$  automatically, and unless  $\|V\|_{2 \rightarrow \infty} \leq Cn^{1/q}$  we will be in the second case.

A corollary of Theorem 2.1 is a subexponential algorithm for SMALL-SET EXPANSION

**Corollary 10.3.** For every  $0.4 > \nu > 0$  there is an  $\exp(n^{1/O(\log(1/\nu))})$  time algorithm that given a graph with the promise that either (i)  $\Phi_G(\delta) \geq 1 - \nu$  or (ii)  $\Phi_G(\delta^2) \leq 0.5$  decides which is the case.

*Proof.* For  $q = O(\log(1/\nu))$  we find from Theorem 2.4 that in case (i),  $\|V_{\geq 0.4}\|_{2 \rightarrow q} \leq 2/\sqrt{\delta}$ , while in case (ii)  $\|V_{\geq 0.4}\|_{2 \rightarrow q} \geq 0.1/\delta^{1-2/q}$ . Thus it suffices to obtain a  $(2/\sqrt{\delta}, 0.1/\delta^{1-2/q})$ -approximation for the  $2 \rightarrow q$  norm to solve the problem, and by Theorem 2.1 this can be achieved in time  $\exp(n^{O(\log(1/\nu))})$  for sufficiently small  $\delta$ .  $\square$

## Conclusions

This work motivates further study of the complexity of approximating hypercontractive norms such as the  $2 \rightarrow 4$  norm. A particularly interesting question is what is the complexity of obtaining a good approximation for the  $2 \rightarrow 4$  norm and what's the relation of this problem to the SMALL-SET EXPANSION problem. Our work leaves possible at least the following three scenarios: **(i)** both these problems can be solved in quasipolynomial time, but not faster, which would mean that the UGC as stated is essentially false but a weaker variant of it is true, **(ii)** both these problems are **NP**-hard to solve (via a reduction with polynomial blowup) meaning that the UGC is true, and **(iii)** the SMALL-SET EXPANSION and UNIQUE GAMES problems are significantly easier than the  $2 \rightarrow 4$  problem with the most extreme case being that the former two problems can be solved in polynomial time and the latter is **NP**-hard and hence cannot be done faster than subexponential time. This last scenario would mean that one can improve on the subexponential algorithm for the  $2 \rightarrow 4$  norm for general instances by using the structure of instances arising from the SMALL-SET EXPANSION reduction of Theorem 2.4 (which indeed seem quite different from the instances arising from the hardness reduction of Theorem 2.5). In any case we hope that further study of the complexity of computing hypercontractive norms can lead to a better understanding of the boundary between hardness and easiness for UNIQUE GAMES and related problems.

**Acknowledgments.** We thank Pablo Parrilo for useful discussions, and the anonymous STOC referees for numerous comments that greatly improved the presentation of this paper. Aram Harrow was funded by NSF grants 0916400, 0829937, 0803478 and DARPA

QuEST contract FA9550-09-1-0044. Jonathan Kelner was partially supported by NSF awards 1111109 and 0843915.

## References

- [ABS10] Sanjeev Arora, Boaz Barak, and David Steurer, *Subexponential algorithms for unique games and related problems*, FOCS, 2010, pp. 563–572. [1](#), [2](#), [4](#), [6](#), [10](#), [42](#)
- [AE07] Andris Ambainis and Joseph Emerson, *Quantum  $t$ -designs:  $t$ -wise independence in the quantum world*, IEEE Conference on Computational Complexity 2007 (2007), 129–140, arXiv:quant-ph/0701126v2. [35](#)
- [ALPTJ10] Radosław Adamczak, Alexander E. Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles*, J. Amer. Math. Soc. **23** (2010), 535–561, arXiv:0903.2323. [24](#)
- [ALPTJ11] ———, *Sharp bounds on the rate of convergence of the empirical covariance matrix*, Comptes Rendus Mathématique **349** (2011), no. 3-4, 195–200, arXiv:1012.0294. [24](#)
- [BaCY11] Fernando G.S.L. Brandão, Matthias Christandl, and Jon Yard, *A quasipolynomial-time algorithm for the quantum separability problem*, Proceedings of the 43rd annual ACM symposium on Theory of computing, STOC '11, 2011, arXiv:1011.2751, pp. 343–352. [3](#), [4](#), [12](#), [40](#), [41](#)
- [BGH<sup>+</sup>11] Boaz Barak, Parikshit Gopalan, Johan Håstad, Raghu Meka, Prasad Raghavendra, and David Steurer, *Making the long code shorter, with applications to the unique games conjecture*, 2011, Manuscript. [6](#), [14](#), [15](#), [16](#), [22](#), [23](#), [49](#)
- [Bis11] Punyashloka Biswal, *Hypercontractivity and its applications*, Manuscript. Available as eprint [arXiv:1101.2913v1](#), 2011. [1](#)
- [BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer, *Rounding semidefinite programming hierarchies via global correlation*, FOCS, 2011, To appear. [arXiv:1104.4680v1](#). [3](#), [6](#)
- [BRSdW11] Harry Buhrman, Oded Regev, Giannicola Scarpa, and Ronald de Wolf, *Near-optimal and explicit bell inequality violations*, Proceedings of the 2011 IEEE 26th Annual Conference on Computational Complexity, CCC '11, 2011, arXiv:1012.5043, pp. 157–166. [42](#)
- [BS10] Salman Beigi and Peter W. Shor, *Approximating the set of separable states using the positive partial transpose test*, J. Math. Phys. **51** (2010), no. 4, 042202, arXiv:0902.1806. [38](#), [41](#)
- [BTN98] Aharon Ben-Tal and Arkadi Nemirovski, *Robust convex optimization*, Mathematics of Operations Research **23** (1998), no. 4, 769–805. [5](#)
- [BV11] Aditya Bhaskara and Aravindan Vijayaraghavan, *Approximating matrix  $p$ -norms*, SODA, 2011, pp. 497–511. [1](#)



- [CFS02] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack, *Unknown quantum states: The quantum de finetti representation*, J. Math. Phys. **43** (2002), no. 9, 4537–4559, arXiv:quant-ph/0104088. [39](#)
- [CKMR07] Matthias Christandl, Robert König, Graeme Mitchison, and Renato Renner, *One-and-a-half quantum de finetti theorems*, Commun. Math. Phys. **273** (2007), 473–498, arXiv:quant-ph/0602130. [39](#), [41](#)
- [CKP00] Fernando Cobos, Thomas Kühn, and Jaak Peetre, *Remarks on symmetries of trilinear forms*, Rev. R. Acad. Cienc. Exact. Fis.Nat. (Esp) **94** (2000), no. 4, 441–449. [32](#)
- [CT10] Eden Chlamtac and Madhur Tulsiani, *Convex relaxations and integrality gaps*, 2010, Chapter in Handbook on Semidefinite, Cone and Polynomial Optimization. [3](#)
- [dKL11] Etienne de Klerk and Monique Laurent, *On the lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems*, SIAM J. on Optimization **21** (2011), no. 3, 824–832. [8](#)
- [dIVKKV05] Wenceslas Fernandez de la Vega, Marek Karpinski, Ravi Kannan, and Santosh Vempala, *Tensor decomposition and approximation schemes for constraint satisfaction problems*, Proceedings of the thirty-seventh annual ACM symposium on Theory of computing, STOC '05, 2005, pp. 747–754. [5](#)
- [DPS04] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri, *Complete family of separability criteria*, Phys. Rev. A **69** (2004), 022308, arXiv:quant-ph/0308032. [3](#), [12](#), [39](#), [41](#)
- [GNN] Francois Le Gall, Shota Nakagawa, and Harumichi Nishimura, *On qma protocols with two short quantum proofs*, arXiv:1108.4306. [34](#), [38](#)
- [Gro75] Leonard Gross, *Logarithmic sobolev inequalities*, Amer. J. Math. **97** (1975), 1061–1083. [1](#)
- [GS11] Venkatesan Guruswami and Ali Kemal Sinop, *Lasserre hierarchy, higher eigenvalues, and approximation schemes for quadratic integer programming with psd objectives*, FOCS, 2011, To appear. [6](#)
- [Gur03] Leonid Gurvits, *Classical deterministic complexity of Edmonds' problem and quantum entanglement*, STOC, 2003, arXiv:quant-ph/0303055, pp. 10–19. [34](#), [38](#)
- [HHH96] Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki, *Separability of mixed states: necessary and sufficient conditions*, Phys. Lett. A **223** (1996), no. 1–2, 1–8, arXiv:quant-ph/9605038. [39](#)
- [HM10] Aram W. Harrow and Ashley Montanaro, *An efficient test for product states, with applications to quantum Merlin-Arthur games*, Proc. 51st Symp. on FOCS, 2010, arXiv:1001.0017. [4](#), [5](#), [11](#), [31](#), [34](#), [35](#), [36](#), [37](#), [38](#)
- [IPZ98] R. Impagliazzo, R. Paturi, and F. Zane, *Which problems have strongly exponential complexity?*, Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on, IEEE, 1998, pp. 653–662. [5](#)

- [Kho02] Subhash Khot, *On the power of unique 2-prover 1-round games*, STOC, 2002, pp. 767–775. [1](#), [4](#)
- [KKMO04] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O’Donnell, *Optimal inapproximability results for Max-Cut and other 2-variable CSPs?*, FOCS, 2004, pp. 146–154. [10](#), [14](#), [15](#), [16](#), [20](#)
- [KM09] Robert Koenig and Graeme Mitchison, *A most compendious and facile quantum de finetti theorem*, J. Math. Phys. **50** (2009), 012105, arXiv:quant-ph/0703210. [41](#)
- [KMN11] Anna R. Karlin, Claire Mathieu, and C. Thach Nguyen, *Integrality gaps of linear and semi-definite programming relaxations for knapsack*, IPCO, 2011, pp. 301–314. [6](#)
- [KNS08] Guy Kindler, Assaf Naor, and Gideon Schechtman, *The UGC hardness threshold of the  $\ell_p$  Grothendieck problem*, SODA, 2008, pp. 64–73. [1](#)
- [KPS10] Subhash Khot, Preyas Popat, and Rishi Saket, *Approximate lasserre integrality gap for unique games*, APPROX-RANDOM, 2010, pp. 298–311. [3](#), [6](#), [8](#), [14](#)
- [KS09] Subhash Khot and Rishi Saket, *SDP integrality gaps with local  $\ell_1$ -embeddability*, FOCS, 2009, pp. 565–574. [6](#), [14](#)
- [KV05] Subhash Khot and Nisheeth K. Vishnoi, *The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into  $\ell_1$* , FOCS, 2005, pp. 53–62. [4](#), [6](#), [9](#), [10](#), [14](#), [15](#), [22](#), [27](#), [49](#)
- [Las01] Jean B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM Journal on Optimization **11** (2001), no. 3, 796–817. [3](#), [6](#), [8](#)
- [Lau03] Monique Laurent, *A comparison of the sherali-adams, lovász-schrijver, and lasserre relaxations for 0-1 programming*, Math. Oper. Res. **28** (2003), no. 3, 470–496. [3](#)
- [Lau09] M. Laurent, *Sums of squares, moment matrices and optimization over polynomials*, Emerging applications of algebraic geometry **149** (2009), 157–270. [3](#), [6](#)
- [Liu07] Yi-Kai. Liu, *The complexity of the consistency and  $N$ -representability problems for quantum states*, Ph.D. thesis, Univ. of California, San Diego, 2007, arXiv:0712.3041. [38](#)
- [LS91] L. Lovász and A. Schrijver, *Cones of matrices and set-functions and 0–1 optimization*, SIAM Journal on Optimization **1** (1991), 166–190. [3](#)
- [MOO05] Elchanan Mossel, Ryan O’Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences invariance and optimality*, FOCS, 2005, pp. 21–30. [10](#), [15](#), [16](#), [19](#)
- [Nes00] Y. Nesterov, *Squared functional systems and optimization problems*, High performance optimization **13** (2000), 405–440. [3](#)

- [NOP09] Miguel Navascues, Masaki Owari, and Martin B. Plenio, *The power of symmetric extensions for entanglement detection*, Phys. Rev. A **80** (2009), 052306, arXiv:0906.2731. 39, 41
- [O'D07] Ryan O'Donnell, *Analysis of boolean functions*, Lecture Notes. Available online at <http://www.cs.cmu.edu/~odonnell/boolean-analysis/>, 2007. 9, 13, 49
- [Par00] Pablo A. Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*, Tech. report, MIT, 2000, Ph.D thesis. 3, 6
- [Par03] Pablo A. Parrilo, *Semidefinite programming relaxations for semialgebraic problems*, Mathematical Programming **96** (2003), 293–320. 3, 6
- [Pis99] G. Pisier, *The volume of convex bodies and banach space geometry*, Cambridge Tracts in Mathematics, Cambridge University Press, 1999. 23, 37
- [RS09] Prasad Raghavendra and David Steurer, *Integrality gaps for strong SDP relaxations of unique games*, FOCS, 2009, pp. 575–585. 3, 6, 8, 14
- [RS10] ———, *Graph expansion and the unique games conjecture*, STOC, 2010, pp. 755–764. 1, 4, 6, 11, 26, 27
- [RST10a] Prasad Raghavendra, David Steurer, and Prasad Tetali, *Approximations for the isoperimetric and spectral profile of graphs and related parameters*, STOC, 2010, pp. 631–640. 4, 27
- [RST10b] Prasad Raghavendra, David Steurer, and Madhur Tulsiani, *Reductions between expansion problems*, Manuscript. Available as eprint [arXiv:1011.2586v1](https://arxiv.org/abs/1011.2586v1), 2010. 4
- [Rya02] Raymond A. Ryan, *Introduction to tensor products of banach spaces*, Springer monographs in mathematics, Springer, 2002. 30
- [SA90] Hanif D. Sherali and Warren P. Adams, *A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems*, SIAM J. Discrete Math. **3** (1990), no. 3, 411–430. MR MR1061981 (91k:90116) 3
- [SC97] Laurent Saloff-Coste, *Lectures on finite markov chains*, Lectures on probability theory and statistics **1665** (1997), 301–413. 1
- [Sch08] Grant Schoenebeck, *Linear level Lasserre lower bounds for certain k-CSPs*, FOCS, 2008, pp. 593–602. 50
- [Sho87] NZ Shor, *An approach to obtaining global extremums in polynomial mathematical programming problems*, Cybernetics and Systems Analysis **23** (1987), no. 5, 695–700. 3
- [Ste56] Elias M. Stein, *Interpolation of linear operators*, Trans. Amer. Math. Soc. **83** (1956), 482–492. 41

- [Ste05] Daureen Steinberg, *Computation of matrix norms with applications to robust optimization*, Master's thesis, Technion, 2005, Available on A. Nemirovski's website <http://www2.isye.gatech.edu/~nemirovs/>. 1
- [Ste10] David Steurer, *Subexponential algorithms for  $d$ -to-1 two-prover games and for certifying almost perfect expansion*, Manuscript, available from the author's website., 2010. 27
- [Tul09] Madhur Tulsiani, *CSP gaps and reductions in the Lasserre hierarchy*, STOC, 2009, pp. 303–312. 10
- [vL09] Charles van Loan, *Future directions in tensor-based computation and modeling*, 2009, Unpublished NSF Workshop Report. 5

## A More facts about pseudo-expectation

In this section we note some additional facts about pseudo-expectation functionals that are useful in this paper.

**Lemma A.1.** *The relation  $P^2 \leq P$  holds if and only if  $0 \leq P \leq 1$ . Furthermore, if  $P^2 \leq P$  and  $0 \leq Q \leq P$ , then  $Q^2 \leq Q$ .*

*Proof.* If  $P \geq 0$ , then  $P \leq 1$  implies  $P^2 \leq P$ . (Multiplying both sides with a sum of squares preserves the order.) On the other hand, suppose  $P^2 \leq P$ . Since  $P^2 \geq 0$ , we also have  $P \geq 0$ . Since  $1 - P = P - P^2 + (1 - P)^2$ , the relation  $P^2 \leq P$  also implies  $P \leq 1$ .

For the second part of the lemma, suppose  $P^2 \leq P$  and  $0 \leq Q \leq P$ . Using the first part of the lemma, we have  $P \leq 1$ . It follows that  $0 \leq Q \leq 1$ , which in turn implies  $Q^2 \leq Q$  (using the other direction of the first part of the lemma).  $\square$

**Fact A.2.** *If  $f$  is a  $d$ -f.r.v. over  $\mathbb{R}^{\mathcal{U}}$  and  $\{P_v\}_{v \in \mathcal{U}}$  are polynomials of degree at most  $k$ , then  $g$  with  $g(v) = P_v(f)$  is a level- $(d/k)$  fictitious random variable over  $\mathbb{R}^{\mathcal{U}}$ . (For a polynomial  $Q$  of degree at most  $d/k$ , the pseudo-expectation is defined as  $\tilde{\mathbb{E}}_g Q(\{g(v)\}_{v \in \mathcal{U}}) := \tilde{\mathbb{E}}_f Q(\{P_v(f)\}_{v \in \mathcal{U}})$ .)*

**Lemma A.3.** *For  $f, g \in L_2(\mathcal{U})$ ,*

$$\langle f, g \rangle \leq \frac{1}{2} \|f\|^2 + \frac{1}{2} \|g\|^2.$$

*Proof.* The right-hand side minus the LHS equals the square polynomial  $\frac{1}{2} \langle f - g, f - g \rangle$   $\square$

**Lemma A.4** (Cauchy-Schwarz inequality). *If  $(f, g)$  is a level-2 fictitious random variable over  $\mathbb{R}^{\mathcal{U}} \times \mathbb{R}^{\mathcal{U}}$ , then*

$$\tilde{\mathbb{E}}_{f,g} \langle f, g \rangle \leq \sqrt{\tilde{\mathbb{E}}_f \|f\|^2} \cdot \sqrt{\tilde{\mathbb{E}}_g \|g\|^2}.$$

*Proof.* Let  $\bar{f} = f / \sqrt{\tilde{\mathbb{E}}_f \|f\|^2}$  and  $\bar{g} = g / \sqrt{\tilde{\mathbb{E}}_g \|g\|^2}$ . Note  $\tilde{\mathbb{E}}_{\bar{f}} \|\bar{f}\|^2 = \tilde{\mathbb{E}}_{\bar{g}} \|\bar{g}\|^2 = 1$ . Since by Lemma A.3,  $\langle \bar{f}, \bar{g} \rangle \leq 1/2 \|\bar{f}\|^2 + 1/2 \|\bar{g}\|^2$ , we can conclude the desired inequality,

$$\tilde{\mathbb{E}}_{f,g} \langle f, g \rangle = \sqrt{\tilde{\mathbb{E}}_f \|f\|^2} \cdot \sqrt{\tilde{\mathbb{E}}_g \|g\|^2} \tilde{\mathbb{E}}_{\bar{f}, \bar{g}} \langle \bar{f}, \bar{g} \rangle \leq \sqrt{\tilde{\mathbb{E}}_f \|f\|^2} \cdot \sqrt{\tilde{\mathbb{E}}_g \|g\|^2} \cdot \underbrace{\left( \frac{1}{2} \tilde{\mathbb{E}}_{\bar{f}} \|\bar{f}\|^2 + \frac{1}{2} \tilde{\mathbb{E}}_{\bar{g}} \|\bar{g}\|^2 \right)}_{=1}. \quad \square$$

**Corollary A.5** (Hölder's inequality). *If  $(f, g)$  is a 4-f.r.v. over  $\mathbb{R}^{\mathcal{U}} \times \mathbb{R}^{\mathcal{U}}$ , then*

$$\tilde{\mathbb{E}}_{f, g} \mathbb{E}_{u \in \mathcal{U}} f(u)g(u)^3 \leq \left( \tilde{\mathbb{E}}_f \|f\|_4^4 \right)^{1/4} \left( \tilde{\mathbb{E}}_g \|g\|_4^4 \right)^{3/4}.$$

*Proof.* Using [Lemma A.4](#) twice, we have

$$\tilde{\mathbb{E}}_{f, g} \mathbb{E}_{u \in \mathcal{U}} f(u)g(u)^3 \leq \left( \tilde{\mathbb{E}}_{f, g} \mathbb{E}_{u \in \mathcal{U}} f(u)^2 g(u)^2 \right)^{1/2} \left( \tilde{\mathbb{E}}_g \|g\|_4^4 \right)^{1/2} \leq \left( \tilde{\mathbb{E}}_f \|f\|_4^4 \right)^{1/4} \left( \tilde{\mathbb{E}}_g \|g\|_4^4 \right)^{3/4}.$$

□

## B Norm bound implies small-set expansion

In this section, we show that an upper bound on  $2 \rightarrow q$  norm of the projector to the top eigenspace of a graph implies that the graph is a small-set expander. This proof appeared elsewhere implicitly [[KV05](#), [O'D07](#)] or explicitly [[BGH<sup>+</sup>11](#)] and is presented here only for completeness. We use the same notation from [Section 8](#). Fix a graph  $G$  (identified with its normalized adjacency matrix), and  $\lambda \in (0, 1)$ , letting  $V_{\geq \lambda}$  denote the subspace spanned by eigenfunctions with eigenvalue at least  $\lambda$ .

If  $p, q$  satisfy  $1/p + 1/q = 1$  then  $\|x\|_p = \max_{y: \|y\|_q \leq 1} |\langle x, y \rangle|$ . Indeed,  $|\langle x, y \rangle| \leq \|x\|_p \|y\|_q$  by Hölder's inequality, and by choosing  $y_i = \text{sign}(x_i) |x_i|^{p-1}$  and normalizing one can see this equality is tight. In particular, for every  $x \in L(\mathcal{U})$ ,  $\|x\|_q = \max_{y: \|y\|_{q/(q-1)} \leq 1} |\langle x, y \rangle|$  and  $\|y\|_{q/(q-1)} = \max_{\|x\|_q \leq 1} |\langle x, y \rangle|$ . As a consequence

$$\|A\|_{2 \rightarrow q} = \max_{\|x\|_2 \leq 1} \|Ax\|_q = \max_{\|x\|_2 \leq 1, \|y\|_{q/(q-1)} \leq 1} |\langle Ax, y \rangle| = \max_{\|y\|_{q/(q-1)} \leq 1} |\langle A^T y, x \rangle| = \|A^T\|_{q/(q-1) \rightarrow 2}$$

Note that if  $A$  is a projection operator,  $A = A^T$ . Thus, part 1 of [Theorem 2.4](#) follows from the following lemma:

**Lemma B.1.** *Let  $G = (V, E)$  be regular graph and  $\lambda \in (0, 1)$ . Then, for every  $S \subseteq V$ ,*

$$\Phi(S) \geq 1 - \lambda - \|V_\lambda\|_{q/(q-1) \rightarrow 2}^2 \mu(S)^{(q-2)/q}$$

*Proof.* Let  $f$  be the characteristic function of  $S$ , and write  $f = f' + f''$  where  $f' \in V_\lambda$  and  $f'' = f - f'$  is the projection to the eigenvectors with value less than  $\lambda$ . Let  $\mu = \mu(S)$ . We know that

$$\Phi(S) = 1 - \langle f, Gf \rangle / \|f\|_2^2 = 1 - \langle f, Gf \rangle / \mu, \quad (\text{B.1})$$

And  $\|f\|_{q/(q-1)} = \left( \mathbb{E} f(x)^{q/(q-1)} \right)^{(q-1)/q} = \mu^{(q-1)/q}$ , meaning that  $\|f'\| \leq \|V_\lambda\|_{q/(q-1) \rightarrow 2} \mu^{(q-1)/q}$ . We now write

$$\begin{aligned} \langle f, Gf \rangle &= \langle f', Gf' \rangle + \langle f'', Gf'' \rangle \leq \|f'\|_2^2 + \lambda \|f''\|_2^2 \leq \|V\|_{q/(q-1) \rightarrow 2}^2 \|f\|_{q/(q-1)}^2 + \lambda \mu \\ &\leq \|V\|_{2 \rightarrow q}^2 \mu^{2(q-1)/q} + \lambda \mu. \end{aligned} \quad (\text{B.2})$$

Plugging this into [\(B.1\)](#) yields the result. □

## C Semidefinite Programming Hierarchies

In this section, we compare different SDP hierarchies and discuss some of their properties.

## C.1 Example of Max Cut

In this section, we compare the SoS hierarchy and Lasserre hierarchy at the example of Max Cut. (We use a formulation of Lasserre's hierarchy similar to the one in [Sch08].) It will turn out that these different formulations are equivalent up to (small) constant factors in the number of levels. We remark that the same proof with syntactic modifications shows that our SoS relaxation of Unique Games is equivalent to the corresponding Lasserre relaxation.

Let  $G$  be a graph (an instance of Max Cut) with vertex set  $V = \{1, \dots, n\}$ . The level- $d$  Lasserre relaxation for  $G$ , denoted  $\text{lass}_d(G)$ , is the following semidefinite program over vectors  $\{v_S\}_{S \subseteq [n], |S| \leq d}$ ,

$$\begin{aligned} \text{lass}_d(G): \text{ maximize} \quad & \sum_{(i,j) \in G} \|v_i - v_j\|^2 \\ \text{subject to} \quad & \langle v_S, v_T \rangle = \langle v_{S'}, v_{T'} \rangle \quad \text{for all sets with } S \Delta T = S' \Delta T', \\ & \|v_\emptyset\|^2 = 1. \end{aligned}$$

The level- $d$  SoS relaxation for  $G$ , denoted  $\text{sos}_d(G)$ , is the following semidefinite program over  $d$ -p.e.f.  $\tilde{\mathbb{E}}$  (and  $d$ -f.r.v.  $x$  over  $\mathbb{R}^V$ ),

$$\begin{aligned} \text{sos}_d(G): \text{ maximize} \quad & \tilde{\mathbb{E}}_x \sum_{(i,j) \in G} (x_i - x_j)^2 \\ \text{subject to} \quad & \tilde{\mathbb{E}}_x (x_i^2 - 1)^2 = 0 \quad \text{for all } i \in V. \end{aligned}$$

**From Lasserre to SoS.** Suppose  $\{v_S\}$  is a solution to  $\text{lass}_d(G)$ . For a polynomial  $P$  over  $\mathbb{R}^V$ , we obtain a multilinear polynomial  $P'$  by successively replacing squares  $x_i^2$  by 1. (In other words, we reduce  $P$  modulo the ideal generated by the polynomials  $x_i^2 - 1$  with  $i \in V$ .) We define a  $d$ -p.e.f.  $\tilde{\mathbb{E}}$  by setting  $\tilde{\mathbb{E}} P = \sum_{|S| \leq d} c_S \langle v_\emptyset, v_S \rangle$ , where  $\{c_S\}_{|S| \leq d}$  are the coefficients of the polynomial  $P' = \sum_{|S| \leq d} c_S \prod_{i \in S} x_i$  obtained by making  $P$  multilinear. The functional  $\tilde{\mathbb{E}}$  is linear (using  $(P + Q)' = P' + Q'$ ) and satisfies the normalization condition. We also have  $\tilde{\mathbb{E}}(x_i^2 - 1)^2 = 0$  since  $(x_i^2 - 1)^2 = 0$  modulo  $x_i^2 - 1$ . Since  $\tilde{\mathbb{E}}_x (x_i - x_j)^2 = \|v_i - v_j\|^2$  for all  $i, j \in V$  (using  $\langle v_\emptyset, v_{ij} \rangle = \langle v_i, v_j \rangle$ ), our solution for  $\text{sos}_d(G)$  has the same objective value as our solution for  $\text{lass}_d(G)$ . It remains to verify positivity. Let  $P^2$  be a polynomial of degree at most  $d$ . We may assume that  $P$  is multilinear, so that  $P = \sum_{|S| \leq d} c_S x_S$ . Therefore  $P^2 = \sum_{S, T} c_S c_T x_S x_T$  and  $\tilde{\mathbb{E}} P^2 = \sum_{S, T} c_S c_T \langle v_\emptyset, v_{S \Delta T} \rangle$ . Using the property  $\langle v_\emptyset, v_{S \Delta T} \rangle = \langle v_S, v_T \rangle$ , we conclude  $\tilde{\mathbb{E}} P^2 = \sum_{S, T} c_S c_T \langle v_S, v_T \rangle = \|\sum_S c_S v_S\|^2 \geq 0$ .

**From SoS to Lasserre.** Let  $\tilde{\mathbb{E}}$  be a solution to  $\text{sos}_d(G)$ . We will construct a solution for  $\text{lass}_{d/2}(G)$  (assuming  $d$  is even). Let  $d' = d/2$ . For  $\alpha \in \mathbb{N}^n$ , let  $x^\alpha$  be the monomial  $\prod_{i \in [n]} x_i^{\alpha_i}$ . The polynomials  $\{x^\alpha\}_{|\alpha| \leq d'}$  form a basis of the space of degree- $d'$  polynomials over  $\mathbb{R}^n$ . Since  $\tilde{\mathbb{E}} P^2 \geq 0$  for all polynomials  $P$  of degree at most  $d'$ , the matrix  $(\tilde{\mathbb{E}} x^\alpha x^\beta)_{|\alpha|, |\beta| \leq d'}$  is positive semidefinite. Hence, there exists vectors  $v_\alpha$  for  $\alpha$  with  $|\alpha| \leq d'$  such that  $\tilde{\mathbb{E}} x^\alpha x^\beta = \langle v_\alpha, v_\beta \rangle$ . We claim that the vectors  $v_\alpha$  with  $\alpha \in \{0, 1\}^n$  and  $|\alpha| \leq d$  form a solution for  $\text{lass}_d(G)$ . The main step is to show that  $\langle v_\alpha, v_\beta \rangle$  depends only on  $\alpha + \beta \pmod 2$ . Since  $\langle v_\alpha, v_\beta \rangle = \tilde{\mathbb{E}} x^{\alpha+\beta}$ , it is enough to show that  $\tilde{\mathbb{E}}$  satisfies  $\tilde{\mathbb{E}} x^\gamma = \tilde{\mathbb{E}} x^{\gamma \pmod 2}$ . Hence, we want to show  $\tilde{\mathbb{E}} x^2 P = \tilde{\mathbb{E}} P$  for all polynomials (with appropriate degree). Indeed, by Lemma 3.5,  $\tilde{\mathbb{E}}(x^2 - 1) \cdot P \leq \sqrt{\tilde{\mathbb{E}}(x^2 - 1)^2} \sqrt{\tilde{\mathbb{E}} P^2} = 0$ .