THEORY OF COMPUTING, Volume 12 (4), 2016, pp. 1–50
www.theoryofcomputing.org

Majority is Stablest: Discrete and SoS

Anindya De* Elchanan Mossel† Joe Neeman‡

Received September 11, 2013; Revised September 9, 2014; Published July 17, 2016

Abstract: The "Majority is Stablest" Theorem has numerous applications in hardness of approximation and social choice theory. We give a new proof of the "Majority is Stablest" Theorem by induction on the dimension of the discrete cube. Unlike the previous proof, it uses neither the "invariance principle" nor Borell's result in Gaussian space. Moreover, the new proof allows us to derive a proof of "Majority is Stablest" in a constant level of the Sum of Squares hierarchy. This implies in particular that the Khot-Vishnoi instance of Max-Cut does not provide a gap instance for the Lasserre hierarchy.

ACM Classification: F.2, G.3

AMS Classification: 68Q17, 60G15, 60D99

Key words and phrases: Majority is Stablest, Sum-of-Squares hierarchy, Gaussian isoperimetry

1 Introduction

The "Majority is Stablest" Theorem [45] affirmed a conjecture in hardness of approximation [30] and in social choice theory [27]. The result has since been extensively used in the two areas. One of the surprising features of the proof of [45] is the crucial use of deep results in Gaussian analysis [8] and an "Invariance Principle" that connects the discrete result to the Gaussian one.

Since the statement of the "Majority is Stablest" Theorem [45] deals with functions on the discrete cube, it is natural to ask (as many have) if there is a "discrete proof" of the statement that "Majority

An extended abstract of this paper appeared in the Proceedings of the 45th ACM Symposium on the Theory of Computing, 2013 [11].

*Supported by Umesh Vazirani's Templeton Foundation Grant 21674.

†Supported by NSF award DMS-1106999, CCF-1320105, DOD ONR grant N000141110140 and Simons Foundation grant 328025.

‡Supported by NSF award DMS-1106999 and DOD ONR grant N00014-11-1-0140, ONR grant N00014-14-1-0823.

 DOI: 10.4086/toc.2016.v012a004

is Stablest." In this paper we answer this question affirmatively and provide a short general proof of the "Majority is Stablest" Theorem. The proof does not rely on Borell's result, nor does it rely on the "Invariance Principle." One advantage of our new proof of "Majority is Stablest" is that it can be turned into a constant level "Sum of Squares" proof [3]. This shows that Khot-Vishnoi instance of MAX-CUT [33] does not provide an integrality gap instance for MAX-CUT in the Lasserre hierarchy [36].

1.1 Functions with low-influence variables

In discrete Fourier analysis, special attention is devoted to functions $f : \{-1,1\}^n \to \{0,1\}$ with low influences. The ith influence of f is defined by

$$\mathrm{Inf}_i(f) = \mathbf{P}[f(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n) \neq f(x_1,\dots,x_{i-1},-x_i,x_{i+1},\dots,x_n)]\,, \tag{1.1}$$

where the probability is taken over the uniform distribution on the discrete cube.

Functions with low influences have played a crucial role in the development of the theory of discrete Fourier analysis. Starting with Kahn, Kalai, and Linial [26], the use of hyper-contractive estimates applied to low-influence variables has become one of the main techniques in discrete Fourier analysis [57, 17].

Of particular interest are functions all of whose influences are low. The work of Friedgut and Kalai [17] shows that such functions have sharp thresholds. Central work in theoretical computer science [9, 14, 52] also pointed to the importance of low-influence functions, including in the context of the "Unique Games Conjecture" [29, 30]. Such functions have also attracted much interest in the theory of social choice, see, e. g., [16, 28].

In the context of voting it is natural to exclude voting schemes that give individual voters too much power. The same is true in the theory of hardness of approximation where a central concept is to distinguish between functions that really depend on many variables versus those which have a strong dependency on a small number of variables, see, e. g., [22, 29, 14]. In this context, it is helpful to recall the definitions of two standard families of Boolean functions. The dictator functions $\{f_i\}_{i\in[n]}$ are defined as $f_i(x) = x_i$. On the other hand the majority function, $\mathrm{MAJ}(x) = \mathrm{SIGN}(\sum x_i)$. The dictator functions are at one end of the spectrum where any of these functions depend on exactly one variable. On the other end is the majority function, where all variables have $o(1)$ influence.

The "Majority is Stablest" Theorem considers the correlation between $f(x)$ and $f(y)$ where $x,y \in \{-1,1\}^n$ are ρ-correlated vectors with $\rho > 0$. Assuming $\mathbf{E}[f] = 1/2$, the function that maximizes $\mathbf{E}[f(x)f(y)]$ is a dictator function. The "Majority is Stablest" theorem states that for functions with low influences the value of $\mathbf{E}[f(x)f(y)]$ cannot be much larger than the corresponding value for the majority function. More formally,

Definition 1.1. For $\rho \in (-1,1)$, the *noise stability* of $f : \{-1,1\}^n \to \mathbb{R}$ at ρ is defined to be

$$\mathrm{Stab}_\rho(f) := \mathbf{E}[f(x)f(y)]\,,$$

when $(x,y) \in \{-1,1\}^n \times \{-1,1\}^n$ is chosen so that $(x_i,y_i) \in \{-1,1\}^2$ are independent random variables with $\mathbf{E}[x_i] = \mathbf{E}[y_i] = 0$ and $\mathbf{E}[x_iy_i] = \rho$.

Theorem 1.2 ("Majority Is Stablest" [45]). *Let $0 \le \rho \le 1$ and $\varepsilon > 0$ be given. Then there exists $\tau > 0$ such that if $f : \{-1,1\}^n \to [0,1]$ satisfies $\mathbf{E}[f] = 1/2$ and $\mathrm{Inf}_i(f) \le \tau$ for all i, then*

$$\mathrm{Stab}_\rho(f) \le 1 - \tfrac{\arccos \rho}{\pi} + \varepsilon\,.$$

By Sheppard's Formula [54], the quantity

$$1 - \frac{\arccos \rho}{\pi}$$

is precisely $\lim_{n\to\infty} \mathrm{Stab}_\rho(\mathrm{Maj}_n)$, where

$$\mathrm{Maj}_n(x_1, \ldots, x_n) = \mathrm{sign}\left(\sum_{i=1}^{n} x_i\right).$$

In particular, the "Majority is Stablest" Theorem shows that no low-influence function can be much more noise stable than the majority function.

We also remark here that Theorem 1.2 readily generalizes to the case when $\mathbf{E}[f] \ne 1/2$, with the right hand side replaced by the corresponding quantity for the shifted majority with the same expectation. This statement of "Majority is Stablest" was conjectured in [30] in the context of hardness of approximation for MAX-CUT. By assuming that Theorem 1.2 holds, the authors showed that it is UNIQUE-GAMES hard[1] to approximate the maximum cut in graphs to within a factor greater than .87856.... This result is optimal, since the efficient algorithm of Goemans and Williamson [19] is guaranteed to find partitions that cut a .87856... fraction of the maximum. A closely related conjecture (for $\rho = -1/3$) was made by Kalai in the context of Arrow's Impossibility Theorem [27]. The results of [45] imply Kalai's conjecture and show that Majority minimizes the probability of Arrow's paradox in ranking 3 alternatives using a balanced ranking function f. See [27, 45] for more details.

The *statement* of Theorem 1.2 deals with Boolean functions, yet *the proof* of [45] crucially relies on Gaussian analysis as (a) it uses a deep result of Borell [8] on noise stability in Gaussian space and (b) it uses the invariance principle developed in [45] that allows to deduce discrete statements from Gaussian statements. This raises the following natural (informal) question:

Question: Is there a "discrete" proof of "Majority is Stablest"?

In other words, does there exist a proof of "Majority is Stablest" not using Borell's result? or any other result in Gaussian space? We note that almost all prior results in discrete Fourier analysis do not use Gaussian results. In particular, the classical hyper-contractive estimates [7, 5] are proved by induction on dimension in the discrete cube. Moreover, most of the results in the area starting from KKL including [26, 57, 17, 9] do not require sophisticated results in Gaussian geometry.

In our main result we provide a positive answer to the question above. Informally we show that

[1] A decision problem P is said to be UNIQUE-GAMES hard, if given an instance G of Unique Games, the problem of deciding whether $\mathrm{VAL}(G) \le \varepsilon$ or $\mathrm{VAL}(G) \ge 1-\varepsilon$ reduces to P for any $\varepsilon > 0$.

Main Result: There is a proof of "Majority is Stablest" by induction on dimension.

Our proof is short and elegant and involves only elementary calculus and hyper-contractivity. The main difficulty in the proof is finding the right statement to prove by induction. The induction statement involves a certain function J, which was recently used in the derivation of a robust version of Borell's result and "Majority is Stablest" [44] using Gaussian techniques and the invariance principle.

In a way, our results here are an analogue of Bobkov's famous inequality in the discrete cube [6]. Bobkov proved by induction a discrete functional inequality that at the limit becomes the Gaussian isoperimetric inequality. Moreover, the functional that Bobkov studied was later used by Bakry and Ledoux [2] to give a semi-group proof of the Gaussian isoperimetric inequality. Moving from isoperimetry to noise sensitivity, the order is reversed: [44] developed a certain functional in order to give a semi-group proof of Borell's result. Here we use the same functional to give a discrete proof by induction.

It is well known that the "Majority is Stablest" Theorem implies Borell's result. Here we show how this can be done by elementary methods only (our proof of Borell's result does not even require hyper-contractivity!). Our proof of Borell's result joins a number of recent proofs of the result including using spherical symmetrization [24], sub-additivity [34] and a semi-group proof [44]. It is the simplest proof of Borell's result using elementary arguments only ([24] uses sophisticated spherical re-arrangement inequalities, [34] only works for sets of measure $1/2$ and certain noise values and [44] requires basic facts on the Orenstein-Uhlenbeck process).

Since it was proved, Theorem 1.2 was generalized a number of times including in [13, 42]. The results and their generalization have been used numerous times in hardness of approximation and social choice theory including in [1, 48, 43, 18]. Our simple proof extends to cover all of the generalizations above. It also enables to prove a Sum-of-Squares version of the statement of "Majority is Stablest," thus answering the main open problem of [49] as we discuss next.

1.2 Sum of Squares proof system

We now discuss an application of our new proof of "Majority is Stablest" to hardness of approximation. To discuss the application, we will first need to introduce the "Sum of Squares" (SoS) proof system. In a nutshell, the SoS proof system is an algebraic proof system where constraints are encoded by polynomial (in)equalities and the deduction rules are specified by a restricted class of polynomial operations. Viewing this proof system as a refutation system for polynomial inequalities, the goal is to show that the given system of constraints is infeasible by using the allowed polynomial operations to arrive at a polynomial constraint which is "obviously" infeasible.

Without further ado, we introduce the following notation: let $X = (X_1, \ldots, X_n)$ be a sequence of variables and let $\mathbb{R}[X]$ be the ring of polynomials on X. For polynomials $p_1, \ldots, p_m \in \mathbb{R}[X]$, let $A = \{p_1 \geq 0, \ldots, p_m \geq 0\}$ be a set of constraints (on $X = (X_1, \ldots, X_n)$). Also, let $\mathbb{M}[X] \subset \mathbb{R}[X]$ be the set of polynomials which can be expressed as sums-of-squares. In other words, $q \in \mathbb{M}[X]$ if and only if $q = r_1^2 + \cdots + r_\ell^2$ for some $r_1, \ldots, r_\ell \in \mathbb{R}[X]$. For $S \subseteq [m]$, we use p_S to denote $\prod_{i \in S} p_i$ with $p_\emptyset = 1$. Now, suppose that there exists $\{q_S\}_{S \subseteq [m]}$ such that for all $S \subseteq [m]$, $q_S \in \mathbb{M}[X]$ and

$$-1 = \sum_{S \subseteq [m]} p_S \cdot q_S \, .$$

Then, it is clear that the constraint set A is infeasible over $\mathbb{R}^n$. The surprisingly powerful theorem of Stengle [56] (and a slightly weaker version proven earlier by Krivine [35]) shows that whenever A is infeasible, such a certificate of infeasibility always exists. This theorem is known as Stengle's Positivstellensatz. Hence, given an infeasible system of constraints A, we can consider a proof system where the proof of infeasibility is given by the set $\{q_S : S \subseteq [m]\}$. We shall refer to this as the Sum of Squares (SoS) proof system. In fact, provided a certain compactness condition holds, the certificate of infeasibility (i. e., the set $\{q_S : S \subseteq [m]\}$) can always be assumed to have $q_S = 0$ for $|S| > 1$; this is due to Putinar [25].

While the results in [35, 56, 25] were well-known in the algebraic geometry community and are intimately tied to Hilbert's seventeenth problem [23], the interest in the theoretical computer science community is relatively new. In the late 1980s, Shor [55] introduced the idea of replacing "positivity constraints" by "sum-of-squares constraints" in optimization problems and also noted that the latter type of constraints (for a fixed degree d) can be enforced using semidefinite programming. Subsequently, this idea appeared in other works in optimization (for example, see Nesterov [46]). The first paper to view Stengle's Positivstellensatz as a proof system for refutation was Grigoriev and Vorobjov [21]. It should be mentioned that an earlier paper by Lombardi, Mnev and Roy [38] also considered the proof theoretic aspects of Positivstellensatz but no attempt was made to quantify the complexity of such proofs. While the degree of the proofs in the SoS proof system can be potentially unbounded, from the point of view of complexity theory, it is also interesting to consider a restricted form of the SoS proof system where one only looks at proofs of refutation where $\max \deg(p_S \cdot q_S) \leq d$. The resulting hierarchy of proof systems where level d corresponds to proofs of refutation of degree at most d, is referred to as the SoS hierarchy.

Besides the obvious motivation from proof complexity, another reason to consider this restricted version is that while one loses completeness (i. e., infeasible constraint sets A may not have a proof of refutation in the degree-d SoS hierarchy for any fixed d), for any fixed d, the restricted proof system is *effective* in the following sense: If the set A has a proof of infeasibility of degree d, then it can be found in time $O(m \cdot n^{O(d)})$ using semidefinite programming. As we have mentioned, while Shor [55] was the first to observe this, the hierarchy was first systematically treated as a tool for optimization by two concurrent but independent works by Jean Lasserre [36] and by Pablo Parrilo (in his Ph. D. thesis) [50]. Strictly speaking, the semidefinite programming hierarchy considered by Lasserre is the dual of the SoS hierarchy considered by Parrilo. Roughly, the dual of the d^{th} level of the Lasserre hierarchy corresponds to the $2d^{th}$ level of the SoS hierarchy. However, in this paper, we will be mainly concerned with the hierarchy at level $d = O(1)$ vs. $d = \omega(1)$ and hence we will interchangeably use the terms "SoS hierarchy" and "Lasserre hierarchy." We remark that while the works of Lasserre and Parrilo were concurrent, Lasserre was the first to connect this hierarchy with the existing lift-and-project hierarchies such as the Lovász-Schrijver hierarchy [40].

Given that the degree-d SoS hierarchy is automatizable, several researchers have tried to understand the limitations of its power. Grigoriev [20] showed linear lower bounds for proofs of refutation of Tseitin tautologies and the *mod* 2 principle. The latter result was essentially rediscovered by Schoenebeck in the Lasserre world independently [53].

Applications to hardness of approximation. While the results of Parillo [50] and Lasserre [36] have been known for more than a decade, there were only a few works in the theoretical computer science community which harnessed the algorithmic power of [50, 36] (see [4, 10]). In fact, for the results which did use Lasserre hierarchy, it was not clear if the full power of Lasserre hierarchy was required, or whether weaker hierarchies, like the one of Lovász and Schrijver [40], would have sufficed.

However, in a recent exciting paper, Barak *et al.* [3] used the degree-8 SoS hierarchy to refute the known integrality gap instances for Unique Games [33, 51, 31]. In other words, there are degree-8 SoS proofs which can be used to certify that the true value of the integrality gap instances is $o(1)$. This is interesting for two reasons. The first is that even after a decade of intense investigation, these integrality gaps remained one of the strongest evidence towards the truth of the Unique Games Conjecture (UGC). Thus, the SoS hierarchy discredits these instances as evidence towards the truth of the UGC. The second reason is that these integrality gaps were known to survive $\Omega((\log\log n)^{1/4})$ rounds of weaker hierarchies like "SDP + Sherali Adams" [51] or "Approximate Lasserre" [32]. Thus, this showed a big gap between the Lasserre/SoS hierarchy and the weaker hierarchies like "SDP+Sherali Adams" or "Approximate Lasserre."

We now mention the main idea behind showing that the degree-8 SoS hierarchy refutes the known integrality gap instances for Unique Games [33, 51, 31]. Analyzing the true optimum of these instances uses tools from analysis like hypercontractivity [7, 5], the KKL theorem [26] etc. Hence, to show that the degree d-SoS hierarchy can refute these instances, one essentially needs to prove SoS versions of these statements in the degree-d SoS hierarchy. Note that so far we have only viewed the SoS as a refutation system, but in fact, as we will see later in the paper, there is an easy extension of the earlier definition, which formalizes the notion of proving a statement in the degree d-SoS hierarchy. In particular, [3] prove SoS versions of results like hypercontractivity, small-set expansion etc.

Extending the results of [3], O'Donnell and Zhou [49] analyze the problems "upward" of unique games (i. e., problems such that Unique Games reduces to these) like MAX-CUT and BALANCED-SEPARATOR. In particular, [49] refutes the integrality gap instances of balanced separator from [12]. Since the key to analyzing the optimum of the BALANCED-SEPARATOR instances in [12] is the KKL theorem [26], the authors provide a proof of the KKL theorem in the degree-4 SoS hierarchy. For MAX-CUT, their results are somewhat less powerful. Again, here they analyze the instances of MAX-CUT from [33]. More precisely, for any $\rho \in (-1,0)$, [33] construct gap-instances of Max-Cut where the true optimum is $\arccos\rho/\pi + o(1)$ whereas the basic SDP-optimum is $(1-\rho)/2+o(1)$. The key to analyzing the true optimum is the "Majority is Stablest" theorem of [45]. Thus, to refute these instances completely—i. e., to show that the true optimum is $\arccos\rho/\pi+o(1)$—requires proving the "Majority is Stablest" theorem in some constant degree-d SoS hierarchy. While O'Donnell and Zhou did not prove that, they did prove the weaker "$2/\pi$" theorem from [30] in some constant degree of the SoS hierarchy. The $2/\pi$ theorem states that for any balanced function whose maximum influence is $o(1)$, the sum of squares of degree-1 Fourier coefficients is bounded by $2/\pi+o(1)$. This implies that the SoS hierarchy can certify that the true optimum is at most $(1/2-\rho/\pi)-(1/2-1/\pi)\rho^3$. They left open the problem of refuting this gap instances optimally, i. e., showing that constant number of rounds of the SoS hierarchy can certify that the true optimum of these gap instances is $\arccos\rho/\pi+o(1)$. In this paper, as the main application of the new proof of "Majority is Stablest," we resolve this problem.

It should be mentioned here that while the new proof of "Majority is Stablest" is more suitable for the SoS hierarchy, several powerful theorems and techniques are needed to achieve this adaptation. For example we use results from approximation theory [39] and a powerful matrix version of Putinar's

Positivstellensatz [37] to prove that a certain polynomial approximation preserves positiveness. We mention here that unlike the previous two papers [3, 49] connecting SoS hierarchy with hardness of approximation, we make essential use of Putinar's Positivstellensatz (essentially, the completeness of the SoS hierarchy). The next theorem shows the power of the SoS hierarchy on the instances of MAX-CUT from [33].

Theorem (SoS-version of Max-Cut)**.** *For every $\delta \in (0,1)$ and $\rho \in (-1,0)$, there exists $d = d(\delta,\rho)$ such that the degree-d SoS hierarchy can certify that the* MAX-CUT *instances from [33] with noise ρ have true optimum less than* $\arccos\rho/\pi + \delta$.

Note that the true optimum for these instances is known to be at least $\arccos\rho/\pi$ and hence the degree-d SoS hierarchy refutes these instances optimally. Also, as a key intermediate step, we establish a SoS version of the "Majority is Stablest" theorem.

Theorem (SoS-version of "Majority is Stablest")**.** *For every $\delta \in (0,1)$ and $\rho \in (-1,0)$, there are constants $c = c(\delta,\rho)$ and $d = d(\delta,\rho)$ such that the following is true: let $0 \le f(x) \le 1$ for all $x \in \{-1,1\}^n$ and $\max_i \mathrm{Inf_i(f)} \le \tau$. There is a degree-d SoS proof of the statement*

$$\mathrm{Stab}_\rho(f) \ge 1 - \arccos\rho/\pi - \delta - c\cdot\tau\,.$$

Our proof can be easily modified to give the analogous statement of "Majority is Stablest" when $\rho \in (0,1)$. Of course, we have to change the direction of the inequality as well as impose the condition that $\mathbf{E}[f] = 1/2$ (this condition is not required when $\rho \in (-1,0)$).

As the reader can see, the theorem is stated very informally. This is because SoS proofs are heavy in notation and it is difficult to express the precise statement without having the proper notation. However, we do remark that the SoS version of MAX-CUT follows easily by composing the proof of refutation of UNIQUE-GAMES instances of [33] (done in [3]) along with the [30] reduction (the proof of soundness of this reduction is the step where we require the SoS version of "Majority is Stablest").

2 Our tensorization theorem

In this section, we will prove our main tensorization inequality on the cube. In subsequent sections, we will use it to give new proofs of the "Majority is Stablest" theorem of Mossel, O'Donnell and Oleszkiewicz [45] and the Gaussian stability inequality of Borell [8]. We begin by recalling a function from [44]: first, let $\Phi : \mathbb{R} \to (0,1)$ denote the cumulative distribution function of a standard normal variable:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\,dy\,.$$

Next, for every $\rho \in [-1,1]$, define $J_\rho : (0,1)^2 \to [0,1]$ as

$$J_\rho(x,y) = \mathbf{Pr}_{X,Y}[X \le \Phi^{-1}(x), Y \le \Phi^{-1}(y)]\,.$$

Here X,Y are jointly normally distributed random variables with the covariance matrix

$$Cov(X,Y) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

We note that [44] proved that in the Gaussian setup where $f,g:\mathbb{R}^n \to [0,1]$, and X,Y are jointly normal random variables with covariance

$$\begin{pmatrix} I_n & \rho I_n \\ \rho I_n & I_n \end{pmatrix},$$

then

$$\mathbf{E} J_\rho(f(X),g(Y)) \le J_\rho(\mathbf{E} f, \mathbf{E} g). \tag{2.1}$$

Applied to indicator functions, this reduces to Borell's inequality [8]. To prove "Majority is Stablest," we would like to show an analogous inequality on the cube, with $f,g:\{-1,1\}^n \to [0,1]$ and X,Y correlated points on $\{-1,1\}^n$. Unfortunately, the inequality (2.1) is not true in this setup, but it is true with some extra error terms on the right hand side. First, let us define the error term:

Definition 2.1. If Ω is a probability space and $f:\Omega^n \to \mathbb{R}$, then for $X \in \Omega$, we define $f_X:\Omega^{n-1} \to \mathbb{R}$ by

$$f_X(X_1,\dots,X_{n-1}) = f(X_1,\dots,X_{n-1},X).$$

Definition 2.2. For a function $f:\Omega \to \mathbb{R}$, define

$$\Delta_1(f) = \mathbf{E}\,|f - \mathbf{E} f|^3.$$

For a function $f:\Omega^n \to \mathbb{R}$, define $\Delta_n(f)$ recursively by

$$\Delta_n(f) = \mathbf{E}_{X_n}[\Delta_{n-1}(f_{X_n})] + \Delta_1(\mathbf{E}[f_{X_n} \mid X_n]),$$

where $\mathbf{E}[f_{X_n} \mid X_n]:\Omega \to \mathbb{R}$ is defined as

$$\mathbf{E}[f_{X_n} \mid X_n](x_n) = \mathbf{E}_{x_1,\dots,x_{n-1}}[f(x_1,\dots,x_n)].$$

Note that the definition of $\Delta_n(f)$ measures the "Lipschitzness" of the function f. For example, if f is L-Lipschitz in the sense that changing any one of the coordinates of input changes the output by at most L, then $\Delta_n(f)$ is bounded by $n \cdot L^3$. Next, we recall the notion of Rényi correlation. This definition will allow us to state a result that applies somewhat more generally than just to correlated points on $\{-1,1\}^n$.

Definition 2.3. Let Ω_1 and Ω_2 be two sets and μ be a probability measure on $\Omega_1 \times \Omega_2$. We say that μ has Rényi correlation at most ρ if for every measurable (w. r. t. μ) $f:\Omega_1 \to \mathbb{R}$ and $g:\Omega_2 \to \mathbb{R}$ with $\mathbf{E}_\mu f = \mathbf{E}_\mu g = 0$,

$$\mathbf{E}_\mu[fg] \le \rho\sqrt{\mathbf{E}_\mu[f^2]\,\mathbf{E}_\mu[g^2]}.$$

For example, suppose that $\Omega_1 = \Omega_2$ and suppose (X,Y) are generated by the following procedure: first choose X according to some distribution ν. Then, with probability ρ, we set $Y = X$, and with probability $1-\rho$, Y is chosen to be an independent sample from ν. If μ is the distribution of (X,Y), then it is easy to check that μ has Rényi correlation ρ.

We prove the following general theorem, which we will later use to derive both Borell's inequality and the "Majority is Stablest" theorem.

Theorem 2.4. *For any $\varepsilon > 0$ and $0 < \rho < 1$, there is $C(\rho) > 0$ such that the following holds. Let μ be a ρ-correlated measure on $\Omega_1 \times \Omega_2$ and let $(X_i,Y_i)_{i=1}^n$ be i. i. d. variables with distribution μ. Then for any measurable functions $f:\Omega_1^n \to [\varepsilon,1-\varepsilon]$ and $g:\Omega_2^n \to [\varepsilon,1-\varepsilon]$,*

$$\mathbf{E} J_\rho(f(X),g(Y)) \le J_\rho(\mathbf{E} f,\mathbf{E} g) + C(\rho)\varepsilon^{-C(\rho)}(\Delta_n(f)+\Delta_n(g)).$$

2.1 The base case

We prove Theorem 2.4 by induction on n. In this section, we will prove the base case $n = 1$:

Claim 2.5. *For any $\varepsilon > 0$ and $0 < \rho < 1$, there is a $C(\rho)$ such that for any two random variables $X, Y \in [\varepsilon, 1-\varepsilon]$ with correlation in $[0, \rho]$,*

$$\mathbf{E} J_\rho(X,Y) \le J_\rho(\mathbf{E}X, \mathbf{E}Y) + C(\rho)\varepsilon^{-C(\rho)}(\mathbf{E}|X - \mathbf{E}X|^3 + \mathbf{E}|Y - \mathbf{E}Y|^3).$$

The proof of Claim 2.5 essentially follows from Taylor's theorem applied to the function J_ρ; the crucial point is that J_ρ satisfies a certain differential equation. Define the matrix $M_{\rho\sigma}(x,y)$ by

$$M_{\rho\sigma}(x,y) = \begin{pmatrix} \frac{\partial^2 J_\rho(x,y)}{\partial x^2} & \sigma \frac{\partial^2 J_\rho(x,y)}{\partial x \partial y} \\ \sigma \frac{\partial^2 J_\rho(x,y)}{\partial x \partial y} & \frac{\partial^2 J_\rho(x,y)}{\partial y^2} \end{pmatrix}.$$

Claim 2.6. *For any $(x,y) \in (0,1)^2$ and $0 \le \sigma \le \rho$, $M_{\rho\sigma}(x,y)$ is a negative semidefinite matrix. Likewise, if $\rho \le \sigma \le 0$, then $M_{\rho\sigma}(x,y)$ is a positive semidefinite matrix.*

We will also use the fact that the third derivatives of J_ρ are bounded (at least, away from the boundary of $[0,1]^2$).

Claim 2.7. *For any $-1 < \rho < 1$, there exists $C(\rho) > 0$ such that for any $i, j \ge 0$, $i + j = 3$,*

$$\left| \frac{\partial^3 J_\rho(x,y)}{\partial x^i \partial y^j} \right| \le C(\rho)(xy(1-x)(1-y))^{-C(\rho)}.$$

Further, the function $C(\rho)$ can be chosen so that it is continuous for $\rho \in (-1,1)$.

Claims 2.6 and 2.7 follow from elementary calculus, and we defer their proofs to the appendix (in fact, Claim 2.6 is implicit in [44] and we include the proof here for the sake of completeness). Now we will use them with Taylor's theorem to prove Claim 2.5.

Proof of Claim 2.5. Fix $\varepsilon > 0$ and $\rho \in (0,1)$. Now let $C(\rho)$ be large enough so that all third derivatives of J_ρ are uniformly bounded by $C(\rho)\varepsilon^{-C(\rho)}$ on the square $[\varepsilon, 1-\varepsilon]^2$ (such a $C(\rho)$ exists by Claim 2.7). Taylor's theorem then implies that for any $a, b, a+x, b+y \in [\varepsilon, 1-\varepsilon]$,

$$\begin{aligned} J_\rho(a+x, b+y) \le J_\rho(a,b) &+ x\frac{\partial J_\rho}{\partial x}(a,b) + y\frac{\partial J_\rho}{\partial y}(a,b) \\ &+ \frac{1}{2}(x\ y) \begin{pmatrix} \frac{\partial^2 J_\rho}{\partial x^2}(a,b) & \frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) \\ \frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) & \frac{\partial^2 J_\rho}{\partial y^2}(a,b) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + C(\rho)\varepsilon^{-C(\rho)}(|x|^3 + |y|^3). \end{aligned} \tag{2.2}$$

Now suppose that X and Y are random variables taking values in $[\varepsilon, 1-\varepsilon]$. If we apply (2.2) with $a = \mathbf{E}X$, $b = \mathbf{E}Y$, $x = X - \mathbf{E}X$, and $y = Y - \mathbf{E}Y$, and then take expectations of both sides, we obtain

$$\mathbf{E}J_\rho(X,Y) \le J_\rho(\mathbf{E}X, \mathbf{E}Y) + \frac{1}{2}\mathbf{E}\left[(\tilde{X}\ \tilde{Y})\begin{pmatrix} \frac{\partial^2 J_\rho}{\partial x^2}(a,b) & \frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) \\ \frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) & \frac{\partial^2 J_\rho}{\partial y^2}(a,b) \end{pmatrix}\begin{pmatrix}\tilde{X} \\ \tilde{Y}\end{pmatrix}\right] + C(\rho)\varepsilon^{-C(\rho)}(\mathbf{E}|\tilde{X}|^3 + \mathbf{E}|\tilde{Y}|^3), \quad (2.3)$$

where $\tilde{X} = X - \mathbf{E}X$ and $\tilde{Y} = Y - \mathbf{E}Y$. Now, if X and Y have correlation $\sigma \in [0,\rho]$ then

$$\mathbf{E}\tilde{X}\tilde{Y} = \sigma\sqrt{\mathbf{E}\tilde{X}^2\mathbf{E}\tilde{Y}^2},$$

and so

$$\mathbf{E}\left[(\tilde{X}\ \tilde{Y})\begin{pmatrix} \frac{\partial^2 J_\rho}{\partial x^2}(a,b) & \frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) \\ \frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) & \frac{\partial^2 J_\rho}{\partial y^2}(a,b) \end{pmatrix}\begin{pmatrix}\tilde{X} \\ \tilde{Y}\end{pmatrix}\right] = (\sigma_X\ \sigma_Y)\begin{pmatrix} \frac{\partial^2 J_\rho}{\partial x^2}(a,b) & \sigma\frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) \\ \sigma\frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) & \frac{\partial^2 J_\rho}{\partial y^2}(a,b) \end{pmatrix}\begin{pmatrix}\sigma_X \\ \sigma_Y\end{pmatrix},$$

where $\sigma_X = \sqrt{\mathbf{E}\tilde{X}^2}$ and $\sigma_Y = \sqrt{\mathbf{E}\tilde{Y}^2}$. By Claim 2.6.

$$(\sigma_X\ \sigma_Y)\begin{pmatrix} \frac{\partial^2 J_\rho}{\partial x^2}(a,b) & \sigma\frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) \\ \sigma\frac{\partial^2 J_\rho}{\partial x \partial y}(a,b) & \frac{\partial^2 J_\rho}{\partial y^2}(a,b) \end{pmatrix}\begin{pmatrix}\sigma_X \\ \sigma_Y\end{pmatrix} = (\sigma_X\ \sigma_Y)M_{\rho\sigma}(a,b)\begin{pmatrix}\sigma_X \\ \sigma_Y\end{pmatrix} \le 0.$$

Applying this to (2.3), we obtain

$$\mathbf{E}J_\rho(X,Y) \le J_\rho(\mathbf{E}X, \mathbf{E}Y) + C(\rho)\varepsilon^{-C(\rho)}(\mathbf{E}|\tilde{X}|^3 + \mathbf{E}|\tilde{Y}|^3).$$ □

2.2 The inductive step

Next, we prove Theorem 2.4 by induction.

Proof of Theorem 2.4. Claim 2.5 proves the base case of the induction. Only the inductive case remains to be proven. Assume that the theorem holds with n replaced by $n-1$. Consider $f : \Omega_1^n \to [\varepsilon, 1-\varepsilon]$ and $g : \Omega_2^n \to [\varepsilon, 1-\varepsilon]$.

Conditioning on (X_n, Y_n) and writing $\tilde{X} = (X_1, \dots, X_{n-1})$, $\tilde{Y} = (Y_1, \dots, Y_{n-1})$, we have

$$\mathbf{E}J_\rho(f(X), g(Y)) = \mathbf{E}_{X_n,Y_n}\mathbf{E}_{\tilde{X},\tilde{Y}}J_\rho(f_{X_n}(\tilde{X}), g_{Y_n}(\tilde{Y})).$$

Applying the inductive hypothesis for $n-1$ conditionally on X_n and Y_n,

$$\mathbf{E}_{\tilde{X},\tilde{Y}}J_\rho(f_{X_n}(\tilde{X}), g_{Y_n}(\tilde{Y})) \le J_\rho(\mathbf{E}[f_{X_n} \mid X_n], \mathbf{E}[g_{Y_n} \mid Y_n]) + C(\rho)\varepsilon^{-C(\rho)}(\Delta_{n-1}(f_{X_n}) + \Delta_{n-1}(f_{Y_n})). \quad (2.4)$$

On the other hand, the base case for $n = 1$ implies that

$$\mathbf{E}_{X_n,Y_n}J_\rho(\mathbf{E}[f_{X_n} \mid X_n], \mathbf{E}[g_{Y_n} \mid Y_n]) \le J_\rho(\mathbf{E}f, \mathbf{E}g) + C(\rho)\varepsilon^{-C(\rho)}(\Delta_1(\mathbf{E}[f_{X_n} \mid X_n]) + \Delta_1(\mathbf{E}[g_{Y_n} \mid Y_n])). \quad (2.5)$$

Taking the expectation of (2.4) over X_n and Y_n and combining it with (2.5), we obtain

$$\begin{aligned}\mathbf{E} J_\rho(f(X), g(Y)) &\le J_\rho(\mathbf{E} f, \mathbf{E} g) + C(\rho)\varepsilon^{-C(\rho)}\big(\mathbf{E}_{X_n}\Delta_{n-1}(f_{X_n}) + \mathbf{E}_{Y_n}\Delta_{n-1}(f_{Y_n})\big) \\ &+ C(\rho)\varepsilon^{-C(\rho)}\big(\Delta_1(\mathbf{E}[f_{X_n} \mid X_n]) + \Delta_1(\mathbf{E}[g_{Y_n} \mid Y_n])\big).\end{aligned}$$

Finally, note that the definition of Δ_n implies that the right-hand side above is just

$$J_\rho(\mathbf{E} f, \mathbf{E} g) + C(\rho)\varepsilon^{-C(\rho)}\big(\Delta_n(f) + \Delta_n(g)\big). \qquad \square$$

3 Borell's inequality

The most interesting special case of Theorem 2.4 is when $\Omega_1 = \Omega_2 = \{-1, 1\}$ and the distributions of X_i, Y_i satisfy $\mathbf{E} X_i = \mathbf{E} Y_i = 0$, $\mathbf{E} X_i Y_i = \rho$. In this section and the next, we will focus on this special case. As we will see in this section, even proving very crude bounds for this case implies Borell's Gaussian stability inequality. First, let us recall the functional version of Borell's inequality that was given in [44].

Theorem 3.1. *Let $\rho \ge 0$ and G_1 and G_2 are Gaussian vectors with joint distribution*

$$(G_1, G_2) \sim \mathcal{N}\left(0, \begin{pmatrix} I_d & \rho I_d \\ \rho I_d & I_d \end{pmatrix}\right).$$

For any measurable $f_1, f_2 : \mathbb{R}^d \to [0, 1]$,

$$\mathbf{E} J_\rho(f_1(G_1), f_2(G_2)) \le J_\rho(\mathbf{E} f_1, \mathbf{E} f_2).$$

We will prove Theorem 3.1 using Theorem 2.4 and a crude bound on $\Delta_n(f)$ (in the next section, we will need a much better bound on $\Delta_n(f)$ to prove that "Majority is Stablest").

Claim 3.2. *For $X \in \{-1, 1\}^n$, define*

$$X^{-i} = (X_1, \dots, X_{i-1}, -X_i, X_{i+1}, \dots, X_n).$$

Then

$$\Delta_n(f) \le \sum_{i=1}^n \frac{\mathbf{E}|f(X) - f(X^{-i})|^3}{8}.$$

Proof. The proof is by induction: the base case is trivial. The inductive step follows by conditioning on X_n:

$$\begin{aligned}\Delta_n(f) &= \mathbf{E}_{X_n}[\Delta_{n-1}(f_{X_n})] + \mathbf{E}_{X_n}|\mathbf{E}[f_{X_n} \mid X_n] - \mathbf{E} f|^3 \\ &\le \sum_{i=1}^{n-1} \frac{\mathbf{E}|f(X) - f(X^{-i})|^3}{8} + \frac{\mathbf{E}_{X_n}|\mathbf{E}[f_{X_n} \mid X_n] - \mathbf{E}[f_{-X_n} \mid X_n]|^3}{8} \\ &\le \sum_{i=1}^{n-1} \frac{\mathbf{E}|f(X) - f(X^{-i})|^3 + \mathbf{E}|f(X) - f(X^{-n})|^3}{8}.\end{aligned}$$

Here the first inequality uses the induction hypothesis while the second inequality uses Jensen's inequality. $\square$

Proof of Theorem 3.1. Let $n = md$ and define

$$G_{1,n} = \frac{1}{\sqrt{m}} \left(\sum_{i=1}^{m} X_i, \sum_{i=m+1}^{2m} X_i, \dots, \sum_{i=(d-1)m+1}^{md} X_i \right).$$

Define $G_{2,n}$ similarly by using Y instead of X. In other words, $G_{1,n}$ and $G_{2,n}$ are vectors obtained by averaging the vectors X and Y over consecutive blocks of size m. Define Z as

$$Z = (X_1, X_{m+1}, \dots, X_{(d-1)m+1}, Y_1, Y_{m+1}, \dots, Y_{(d-1)m+1}).$$

Observe that $(G_{1,n}, G_{2,n})$ is distributed as sum of m independent copies of Z scaled by $1/\sqrt{m}$. Applying the Lindeberg-Feller central limit theorem [15], we obtain $(G_{1,n}, G_{2,n}) \xrightarrow{d} (G_1, G_2)$ as $m \to \infty$.

Suppose first that f_1 and f_2 are L-Lipschitz functions taking values in $[\varepsilon, 1-\varepsilon]$. Define $g_1, g_2 : \{-1,1\}^n \to \mathbb{R}$ as

$$g_j(z) = f_j \left(\frac{1}{\sqrt{m}} \left(\sum_{i=1}^{m} z_i, \sum_{i=m+1}^{2m} z_i, \dots, \sum_{i=(d-1)m+1}^{md} z_i \right) \right).$$

By Theorem 2.4,

$$\mathbf{E} J_\rho(g_1(X), g_2(Y)) \le J_\rho(\mathbf{E} g_1, \mathbf{E} g_2) + C(\rho)\varepsilon^{-C(\rho)} (\Delta_n(g_1) + \Delta_n(g_2)). \tag{3.1}$$

Since f_i is L-Lipschitz,

$$|g_i(X) - g_i(X^{-j})| \le \frac{2L}{\sqrt{m}},$$

for every j, and so Claim 3.2 implies that

$$\Delta_n(g_i) \le \frac{L^3 n}{m^{3/2}} = \frac{L^3 d}{\sqrt{m}}.$$

Applying this to (3.1),

$$\mathbf{E} J_\rho(g_1(X), g_2(Y)) \le J_\rho(\mathbf{E} g_1, \mathbf{E} g_2) + C(\rho)\varepsilon^{-C(\rho)} \frac{2L^3 d}{\sqrt{m}},$$

and so the definition of g_i implies

$$\mathbf{E} J_\rho(f_1(G_{1,n}), f_2(G_{2,n})) \le J_\rho(\mathbf{E} f_1(G_{1,n}), \mathbf{E} f_2(G_{2,n})) + C(\rho)\varepsilon^{-C(\rho)} \frac{2L^3 d}{\sqrt{m}}.$$

Taking $m \to \infty$, the multivariate central limit theorem implies that

$$\mathbf{E} J_\rho(f_1(G_1), f_2(G_2)) \le J_\rho(\mathbf{E} f_1(G_1), \mathbf{E} f_2(G_2)). \tag{3.2}$$

This establishes the theorem for functions f_1 and f_2 which are Lipschitz and take values in $[\varepsilon, 1-\varepsilon]$. But any measurable $f_1, f_2 : \mathbb{R}^d \to [0,1]$ can be approximated (say in $L^p(\mathbb{R}^d, \gamma_d)$) by Lipschitz functions with values in $[\varepsilon, 1-\varepsilon]$. Since neither the Lipschitz constant nor ε appears in (3.2), the general statement of the theorem follows from the dominated convergence theorem. □

4 "Majority is Stablest"

By giving a bound on $\Delta_n(f)$ that is better than Claim 3.2, we can derive the "Majority is Stablest" theorem from Theorem 2.4. Indeed, we can express $\Delta_n(f)$ in terms of the Fourier coefficients of f, and we can bound $\Delta_n(f)$ in terms of the maximum influence of any variable in f. For this, we will introduce some very basic Fourier analytic preliminaries below.

Fourier analysis

Consider the domain $\{-1,1\}^n$ equipped with the uniform measure. We start by defining the "character" functions, i. e., for every $S \subseteq [n]$, define $\chi_S(x) : \{-1,1\}^n \to \mathbb{R}$ as $\chi_S(x) = \prod_{i \in S} x_i$. Now, every function $f : \{-1,1\}^n \to \mathbb{R}$ can be expressed as

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x) \quad \text{where} \quad \widehat{f}(S) = \mathop{\mathbf{E}}_{x \in \{-1,1\}^n} [f(x) \cdot \chi_S(x)] \,.$$

The coefficients $\widehat{f}(S)$ are called the Fourier coefficients of f and the expansion of f in terms of $\widehat{f}(S)$ is called the Fourier expansion of f. It is easy to show that

$$\sum_{S \subseteq [n]} \widehat{f}^2(S) = \mathop{\mathbf{E}}_{x \in \{-1,1\}^n} [f^2(x)] \,.$$

This is known in the literature as Parseval's identity. For the sake of conciseness, we use $\widehat{f}(i)$ to denote $\widehat{f}(\{i\})$ for $1 \le i \le n$. Similarly, for any $\rho \in [-1,1]$ and $x \in \{-1,1\}^n$, we define $y \sim_\rho x$ as the distribution over $\{-1,1\}^n$ where every bit of y is independent and $\mathbf{E}[x_i y_i] = \rho$. This immediately lets us define the noise operator T_ρ as follows: for any function $f : \{-1,1\}^n \to \mathbb{R}$,

$$T_\rho f(x) = \mathop{\mathbf{E}}_{y \sim_\rho x} [f(y)] \,.$$

The effect of the noise operator T_ρ is particularly simple to describe on the Fourier spectrum: $\widehat{T_\rho f}(S) = \rho^{|S|}\widehat{f}(S)$. The reader is referred to the excellent set of lecture notes by Ryan O'Donnell [47] for an extensive reference on this topic.

It is also important to remark here that while we prove the "Majority is Stablest" theorem for the hypercube with the uniform measure, one can easily derive analogues of this theorem for more general product spaces by extending our machinery. Instead of using the Fourier expansion of the function, one has to use the Efron-Stein decomposition (see the lecture notes by Mossel [41] for an extensive reference on the Efron-Stein decomposition). All the statements that we prove here have analogues in the Efron-Stein world. We leave it to the expert reader to fill in the details.

We start by extending the notation of Definition 2.1:

Definition 4.1. For disjoint sets $S, T \subset [n]$, and elements $x \in \{-1,1\}^S, y \in \{-1,1\}^T$, we write $x \cdot y$ for their concatenation in $\{-1,1\}^{S \cup T}$.

For a function $f : \{-1,1\}^n \to \mathbb{R}$, a set $S \subset [n]$, and an element $x \in \{-1,1\}^S$, we define $f_x : \{-1,1\}^{[n] \setminus S} \to \mathbb{R}$ by $f_x(y) = f(x \cdot y)$.

Our first observation is that $\Delta_n(f)$ can be written in terms of Fourier coefficients of random restrictions of f.

Claim 4.2. *Let S_i be defined as $S_i = \{i+1,\dots,n\}$, then*

$$\Delta_n(f) = \sum_{i=1}^{n} \mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{f_X}(i)|^3 .$$

Proof. The proof is by induction. The base case is just the fact that for a function $f:\{-1,1\}\to\mathbb{R}$,

$$|\widehat{f}(1)|^3 = \left|\frac{f(1)-f(-1)}{2}\right|^3 = (\mathbf{E}\,|f-\mathbf{E}\,f|)^3 .$$

For the inductive step, we have

$$\begin{aligned}\Delta_n(f) &= \mathbf{E}_{X_n}[\Delta_{n-1}(f_{X_n})] + \Delta_1(\mathbf{E}[f_{X_n} \mid X_n]) = \mathbf{E}_{X_n}[\Delta_{n-1}(f_{X_n})] + |\widehat{f}(n)|^3 \\ &= \mathbf{E}_{X_n}\left[\sum_{i=1}^{n-1} \mathbf{E}_{X_{i+1},\dots,X_{n-1}} \left| \widehat{f_X}(i)|^3\right.\right] + |\widehat{f}(n)|^3 = \sum_{i=1}^{n} \mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{f_X}(i)|^3 .\end{aligned}$$

The first equality uses the definition of $\Delta_n(f)$ and the third equality uses the induction hypothesis. The second equality follows simply from the definition of $\mathbf{E}[f_{X_n} \mid X_n]$. $\square$

In order to control the Fourier coefficients of restrictions of f, we first express them in terms of the Fourier coefficients of f:

Claim 4.3. *For any disjoint S and U and any $x\in\{-1,1\}^S$,*

$$\widehat{f_x}(U) = \sum_{T\subset S} \chi_T(x)\widehat{f}(T\cup U) .$$

Proof. Fix S and x. Let $g:\{-1,1\}^n\to\mathbb{R}$ be the function such that $g(y)=1$ when $y_i = x_i$ for all $i\in S$, and $g(y)=0$ otherwise. It is easy to check that the Fourier expansion of g is

$$g(y) = 2^{-|S|}\sum_{T\subset S}\chi_T(x)\chi_T(y) .$$

Then

$$\begin{aligned}\widehat{f_x}(U) &= \mathbf{E}_{X_{[n]\setminus S}}[f_x(X_{[n]\setminus S})\cdot\chi_U(X_{[n]\setminus S})] = 2^{|S|}\,\mathbf{E}_X[f(X)g(X)\chi_U(X)] \\ &= \mathbf{E}_X\Big[f(X)\sum_{T\subset S}\chi_T(x)\chi_T(X)\chi_U(X)\Big] = \sum_{T\subset S}\chi_T(x)\widehat{f}(T\cup U) .\end{aligned}$$

$\square$

In particular, the identity of Claim 4.3 allows us to compute second moments of $\widehat{f_X}$:

Claim 4.4. *For any function $f:\{-1,1\}^n\to\mathbb{R}$, any $x\in\{-1,1\}^S$ and any $i\in U\subset[n]$,*

$$\mathbf{E}_{X\in\{-1,1\}^S}|\widehat{f_X}(U)|^2 \le \mathrm{Inf}_i(f) .$$

Moreover, if $S_i = \{i+1,\dots,n\}$ then

$$\sum_{i=1}^{n}\mathbf{E}_{X\in\{-1,1\}^{S_i}}|\widehat{f_X}(i)|^2 = \mathrm{Var}(f) .$$

Proof. In view of Claim 4.3, we can write

$$|\widehat{f_X}(U)|^2 = \sum_{T,T' \subset S} \chi_T(X)\chi_{T'}(X)\widehat{f}(T \cup U)\widehat{f}(T' \cup U).$$

When we take the expectation with respect to X_S, $\mathbf{E}\,\chi_T(X)\chi_{T'}(X) = \delta_{T,T'}$ (where $\delta_{T,T'}$ denotes the indicator for the event $T = T'$) and so the cross-terms vanish:

$$\mathbf{E}_{X_S} |\widehat{f_X}(U)|^2 = \sum_{T \subset S} \widehat{f}^2(T \cup U). \tag{4.1}$$

Since $\mathrm{Inf}_i(f) = \sum_{T \ni i} \widehat{f}^2(T)$, the first part of the claim follows.

For the second part,

$$\sum_{i=1}^{n} \mathbf{E}_{X \in \{-1,1\}^{S_i}} |\widehat{f_X}(i)|^2 = \sum_{i=1}^{n} \sum_{T \subset S_i} \widehat{f}^2(T \cup \{i\}) = \sum_{U \subset [n], U \neq \emptyset} \widehat{f}^2(U), \tag{4.2}$$

where the last equality used the fact that every non-empty $U \subset [n]$ can be written uniquely in the form $T \cup \{i\}$ for some $T \subset \{i+1,\dots,n\}$. But of course the right-hand side of (4.2) is just $\mathrm{Var}(f)$. □

Next, we will consider $\widehat{f}_x(n-i)$ as a polynomial in x and apply hypercontractivity to the right hand side of Claim 4.2. First, note that T_σ commutes (up to a multiplicative factor) with restriction:

Claim 4.5. *For any $0 < \sigma < 1$, if $S, U \subset [n]$ are disjoint then, as polynomials in $x = (x_i)_{i \in S}$,*

$$\widehat{(T_\sigma f)_x}(U) = \sigma^{|U|} T_\sigma(\widehat{f}_x(U)).$$

Proof. By Claim 4.3,

$$\widehat{f}_x(U) = \sum_{T \subset S} \chi_T(x)\widehat{f}(T \cup U).$$

Since $\widehat{T_\sigma f}(T \cup U) = \sigma^{|T|+|U|}\widehat{f}(S)$ and $T_\sigma \chi_T(x) = \sigma^{|T|}\chi_T$, it follows that

$$\widehat{(T_\sigma f)_x}(U) = \sum_{T \subset S} \sigma^{|T|+|U|}\chi_T(x)\widehat{f}(T \cup U) = \sigma^{|U|} T_\sigma(\widehat{f}_x(U)).$$ □

At this point, we recall the well-known Bonami-Beckner hypercontractive inequality:

Theorem 4.6 ([7, 5])**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $1 \le q \le p$. Then, for any $\rho \le \sqrt{(q-1)/(p-1)}$,*

$$\|T_\rho f\|_p \le \|f\|_q.$$

Now, set $q = 2$ and $p = 1 + \sigma^{-2}$ for some $0 < \sigma < 1$. Let S and U be disjoint subsets of $[n]$. Defining the function $g : \{-1,1\}^S \to \mathbb{R}$ as $g(x) = \widehat{f}_x(U)$, we apply Theorem 4.6 to get

$$\mathbf{E}_{X \in \{-1,1\}^S} |T_\sigma(\widehat{f_X}(U))|^p \le (\mathbf{E}_{X \in \{-1,1\}^S} |\widehat{f_X}(U)|^2)^{p/2}.$$

Claim 4.5 then implies that

$$\frac{1}{\sigma^{p|U|}} \cdot \mathbf{E}_{X\in\{-1,1\}^S} |\widehat{(T_\sigma f)_X}(U)|^p \le (\mathbf{E}_{X\in\{-1,1\}^S} |\widehat{f}_X(U)|^2)^{p/2},$$

and since $0 \le \sigma \le 1$, we may remove it from the left hand side:

$$\mathbf{E}_{X\in\{-1,1\}^S} |\widehat{(T_\sigma f)_X}(U)|^p \le (\mathbf{E}_{X\in\{-1,1\}^S} |\widehat{f}_X(U)|^2)^{p/2}.$$

Specializing this to $S = S_i = \{i+1,\dots,n\}$ and $U = U_i = \{i\}$ (for any $1 \le i \le n$), we obtain

$$\mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{(T_\sigma f)_X}(i)|^p \le (\mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{f}_X(i)|^2)^{p/2}.$$

Applying Claim 4.4 to the right hand side, we get

$$\mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{(T_\sigma f)_X}(i)|^p \le \mathrm{Inf}_i(f)^{\frac{p-2}{2}} \cdot (\mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{f}_X(i)|^2).$$

Summing over from $i = 1$ to n, we get

$$\sum_{i=1}^n \mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{(T_\sigma f)_X}(i)|^p \le \sum_{i=1}^n \mathrm{Inf}_i(f)^{\frac{p-2}{2}} \cdot (\mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{f}_X(i)|^2) \le \left(\max_i \mathrm{Inf}_i(f)\right)^{\frac{p-2}{2}} \cdot \mathrm{Var}(f), \quad (4.3)$$

where the last inequality uses the fact that

$$\sum_{i=1}^n (\mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{f}_X(i)|^2) = \sum_{S\neq\phi} \widehat{f}^2(S) = \mathrm{Var}(f).$$

As a consequence, we have the following claim:

Claim 4.7. *If* $1+\sigma^{-2} \le 3$ *then*

$$\Delta_n(T_\sigma f) \le (\max_i \mathrm{Inf}_i(f))^{\frac{1-\sigma^2}{2\sigma^2}}.$$

Proof. Using $Range(f) \subseteq [-1,1]$, we get that for any $X \in \{-1,1\}^{S_i}$ and $|\widehat{(T_\sigma f)_X}(i)| \le 1$. As a consequence, using Claim 4.2, we get that for $2 \le p \le 3$,

$$\Delta_n(T_\sigma f) \le \left(\sum_{i=1}^n \mathbf{E}_{X\in\{-1,1\}^{S_i}} |\widehat{(T_\sigma f)_X}(i)|^p\right).$$

Finally, using (4.3), we get the desired conclusion. □

Now we are ready to prove the "Majority is Stablest" theorem. For this, we define $\mathbb{S}_\rho(f)$ as

$$\mathbb{S}_\rho(f) = \mathbf{E}_{x\in\{-1,1\}^n, y\sim_\rho x}[f(x)f(y)].$$

Theorem 4.8. *For any* $0 < \rho < 1$*, there are constants* $0 < c(\rho), C(\rho) < \infty$ *such that for any function* $f : \{-1,1\}^n \to [0,1]$ *with* $\max_i \mathrm{Inf}_i(f) \le \tau$*,*

$$\mathbb{S}_\rho(f) \le J_\rho(\mathbf{E} f, \mathbf{E} f) + C(\rho)\frac{\log\log(1/\tau)}{\log(1/\tau)}.$$

As remarked earlier, our proof extends to the generalizations of Theorem 4.8 such as those presented by [42]. The extension of the proof uses the Efron-Stein decomposition instead of the Fourier decomposition. The only difference is that the hyper-contractivity parameter will now depend on the underlying space. See [42] for more details.

Proof. Suppose $f : \{-1,1\}^n \to [\varepsilon, 1-\varepsilon]$ satisfies $\max_i \mathrm{Inf}_i(f) \le \tau$, and let X,Y be uniformly random elements of $\{-1,1\}^n$ with $\mathbf{E} X_i Y_i = \rho$. By setting $\sigma = 1-\eta$ for sufficiently small $\eta > 0$, Claim 4.7 implies that

$$\Delta_n(T_{1-\eta} f) \le \tau^{\eta/2}\,.$$

Note that the range of $T_{1-\eta} f$ belongs to $[\varepsilon, 1-\varepsilon]$ because the range of f does. Hence, Theorem 2.4 applied to $T_{1-\eta} f$ implies that

$$\mathbf{E} J_\rho(T_{1-\eta} f(X), T_{1-\eta} f(Y)) \le J_\rho(\mathbf{E} T_{1-\eta} f, \mathbf{E} T_{1-\eta} f) + \Delta_n(T_{1-\eta} f) \le J_\rho(\mathbf{E} f, \mathbf{E} f) + C(\rho)\varepsilon^{-C(\rho)}\tau^{c\eta}\,.$$

Since $J_\rho(x,y) \ge xy$, it follows that

$$\mathbb{S}_{\rho(1-\eta)^2}(f) = \mathbf{E} T_{1-\eta} f(X) T_{1-\eta} f(Y) \le J_\rho(\mathbf{E} f, \mathbf{E} f) + C\varepsilon^{-C(\rho)}\tau^{c\eta}\,.$$

This inequality holds for any $0 < \rho < 1$; hence we can replace $\rho(1-\eta)^2$ by ρ to obtain

$$\mathbb{S}_\rho(f) = \mathbf{E} T_{1-\eta} f(X) T_{1-\eta} f(Y) \le J_{\rho(1-\eta)^{-2}}(\mathbf{E} f, \mathbf{E} f) + C\varepsilon^{-C(\rho)}\tau^{c\eta} \tag{4.4}$$

for any $\rho \le (1-\eta)^2$.

Now, (4.4) holds for any $f : \{-1,1\}^n \to [\varepsilon, 1-\varepsilon]$. For a function f taking values in $[-1,1]$, let f^ε be f truncated to $[\varepsilon, 1-\varepsilon]$. Since $|\mathbf{E} f^\varepsilon - \mathbf{E} f| \le \varepsilon$ and (by the proof of Claim 2.6)

$$\frac{\partial J_\rho(x,y)}{\partial x} \le 1$$

for any ρ,

$$J_{\rho(1-\eta)^{-2}}(\mathbf{E} f^\varepsilon, \mathbf{E} f^\varepsilon) \le J_{\rho(1-\eta)^{-2}}(\mathbf{E} f, \mathbf{E} f) + 2\varepsilon\,.$$

On the other hand, $|f - f^\varepsilon| \le \varepsilon$ and so

$$\mathbb{S}_\rho(f) = \mathbf{E} f(X) f(Y) \ge \mathbb{S}_\rho(f^\varepsilon) - 2\varepsilon\,.$$

Thus, (4.4) applied to f^ε implies that for any $\rho \le (1-\eta)^2$ and any $\varepsilon > 0$,

$$\mathbb{S}_\rho(f) \le J_{\rho(1-\eta)^{-2}}(\mathbf{E} f, \mathbf{E} f) + 2\varepsilon + C\varepsilon^{-C(\rho)}\tau^{c\eta}\,.$$

If we set $\varepsilon = \tau^{c\eta/(2C(\rho))}$ then

$$\mathbb{S}_\rho(f) \le J_{\rho(1-\eta)^{-2}}(\mathbf{E} f, \mathbf{E} f) + C\tau^{c(\rho)\eta}\,.$$

Finally, some calculus on J_ρ (see Claim A.1) shows that

$$\left|\frac{\partial J_\rho(x,y)}{\partial \rho}\right| \le \left(\sqrt{1-\rho^2}\right)^{-3/2}$$

for any x,y; hence

$$\mathbb{S}_\rho(f) \le J_{\rho(1-\eta)^{-2}}(\mathbf{E} f, \mathbf{E} f) + \frac{(1-\eta)^{-2}-1}{(1-\rho^2)^{3/2}} + C\tau^{c(\rho)\eta} \le J_\rho(\mathbf{E} f, \mathbf{E} f) + C(\rho)(\eta + \tau^{c(\rho)\eta}).$$

Choosing

$$\eta = C(\rho)\frac{\log\log(1/\tau)}{\log(1/\tau)}$$

completes the proof as long as $\rho \le (1-\eta)^2$. However, we can trivially make the theorem true for $(1-\eta)^2 \le \rho$ by choosing $C(\rho)$ and $c(\rho)$ appropriately. □

5 Sum of Squares hierarchy

In this section, we formally give an introduction to the Sum of Squares (hereafter abbreviated as SoS) hierarchy. To define the SoS hierarchy, let $X = (x_1, \ldots, x_n)$ and let $\mathbb{R}[X]$ be the ring of real polynomials over these variables. We also let $\mathbb{R}_{\le d}[X]$ denote the subset of $\mathbb{R}[X]$ consisting of polynomials of total degree bounded by d. As before, let $\mathbb{M}[X] \subset \mathbb{R}[X]$ be the set of polynomials which can be expressed as sums-of-squares. For polynomials $p_1, \ldots, p_m$ and $q_1, \ldots, q_\ell$, consider two sets of constraints,

- $A_e = \{p_1(X) = 0, p_2(X) = 0, \ldots, p_m(X) = 0\}$;
- $A_g = \{q_1(X) \ge 0, q_2(X) \ge 0, \ldots, q_\ell(X) \ge 0\}$.

Before we go ahead, we define the set $\mathcal{M}_{n,d}[X]$ as the set of monomials over $x_1, \ldots, x_n$ of degree bounded by d. Also, let $\mathbb{M}_{\le d}[X]$ denote the subset of $\mathbb{M}[X]$ of polynomials of degree bounded by d. Further, if $A = A_e \cup A_g$, define $\mathbb{V}(A) = \{X \in \mathbb{R}^n : A \text{ holds on } X\}$. We next define the (degree d) closure of these constraints:

$$\mathcal{C}_d(A_e) = \{p_s(X) \cdot p(X) : s \in [m],\, p(X) \in \mathcal{M}_{n,d}[X] \text{ and } \deg(p) + \deg(p_s) \le d\};$$
$$\mathcal{C}_d(A_g) = \left\{\prod_{i=1}^m q_i^{a_i}(X) : a_1, \ldots, a_m \in \mathbb{Z}^+ \text{ and } \sum_{i=1}^m a_i \cdot \deg(q_i) \le d\right\}.$$

Note that $\mathcal{C}_d(A_g)$ includes $1 \in \mathbb{R}$. Also, it is obvious that the following constraints are implied by $A_e \cup A_g$: For $p(X) \in \mathcal{C}_d(A_e)$, $p(X) = 0$ and for $q(X) \in \mathcal{C}_d(A_g)$, $q(X) \ge 0$. We also observe that the sets $\mathcal{C}_d(A_e)$ and $\mathcal{C}_d(A_g)$ can be computed from A_e and A_g in time $n^{O(d)} \cdot \max\{\ell, m\}$.

Definition 5.1. For the constraint set $A = A_e \cup A_g$ defined above and $h(X) \in \mathbb{R}[X]$, we say $A \vdash_d h(X) \ge 0$ if and only if

$$h(X) = \sum_{p(X) \in \mathcal{C}_d(A_e)} \alpha_p \cdot p(X) + \sum_{q(X) \in \mathcal{C}_d(A_g)} r_q(X) \cdot q(X),$$

where $\alpha_p \in \mathbb{R}$, $r_q \in \mathbb{M}[X]$ and for all $q(X) \in \mathcal{C}_d(A_g)$, $\deg(r_q) + \deg(q) \le d$. In this case, we say that A degree-d SoS proves $h(X) \ge 0$.

For the constraint set A, we say that $A \vdash_d -1 \geq 0$ if and only if there exists

$$-1 = \sum_{p(X) \in \mathcal{C}_d(A_e)} \alpha_p \cdot p(X) + \sum_{q(X) \in \mathcal{C}_d(A_g)} r_q(X) \cdot q(X)$$

with the same constraints on α_p and r_q as above. In this case, we say that there is a degree-d SoS refutation of the constraint set A.

Note that we are adopting the same notation as in [49]. The reason we are interested in Definition 5.1 is because one can efficiently decide if $A \vdash_d -1 \geq 0$ using semidefinite programming. This is because deciding if $A \vdash_d -1 \geq 0$ is equivalent to refuting the existence of a map $\widetilde{E} : \mathbb{R}_{\leq d}[X] \to \mathbb{R}$ satisfying the following conditions (see [50] for more details):

- $\widetilde{E}(1) = 1$.
- It is a linear map, i. e., for every $g, h \in \mathbb{R}_{\leq d}[X]$ and $\alpha, \beta \in \mathbb{R}$, $\widetilde{E}(\alpha g + \beta h) = \alpha \widetilde{E}(g) + \beta \widetilde{E}(h)$.
- For every $h \in \mathcal{C}_d(A_e)$, $\widetilde{E}(h) = 0$.
- For every $h \in \mathcal{C}_d(A_g)$ and $g \in \mathbb{M}_{\leq d}[X]$, such that $\deg(g \cdot h) \leq d$, $\widetilde{E}(g \cdot h) \geq 0$.

A map $\widetilde{E}$ which satisfies all the above constraints is called a degree-d SoS consistent map for the constraint set $A = A_e \cup A_g$. Lasserre [36] and Parillo [50] have shown that using semidefinite programming, it is possible to decide the feasibility of such a map $\widetilde{E}$ in time $m \cdot n^{O(d)}$. In fact, if there exists such a map $\widetilde{E}$, then the algorithm outputs one in the same time. It is important to mention that since the domain of $\widetilde{E}$ is not finite, it is not obvious what one means by outputting the map. To see why this makes sense, note that $\widetilde{E}$ is a linear map and hence it suffices to give to specify $\widetilde{E}$ on the set $\mathcal{M}_{n,d}[X]$. We also remark here that the notion of finding a mapping $\tilde{E}$ is close to the viewpoint taken by Barak *et al.* [3].

To get started with SoS proof systems, we state a few facts (which are very easy to prove):

Fact 5.2.

- *If* $A \vdash_d p \geq 0$ *and* $A' \vdash_{d'} q \geq 0$*, then* $A \cup A' \vdash_{\max\{d,d'\}} p + q \geq 0$.
- *If* $A \vdash_d p \geq 0$ *and* $A \vdash_{d'} q \geq 0$*, then* $A \vdash_{d+d'} p \cdot q \geq 0$.
- *If* $A \vdash_d \{p_1 \geq 0, p_2 \geq 0, \ldots, p_m \geq 0\}$ *and* $\{p_1 \geq 0, p_2 \geq 0, \ldots, p_m \geq 0\} \vdash_{d'} q \geq 0$*,* $A \vdash_{d \cdot d'} q \geq 0$.

We next prove several other propositions in the SoS proof system. It will be helpful for the uninitiated reader to look at these proofs in order to get familiar with the SoS proof system. For rest of the paper, we set the following convention for indeterminates appearing in SoS proofs: capital letters X, Y and Z will be used to denote a sequence of indeterminates (i. e., $X = (x_1, \ldots, x_n)$) while small letters x, y and z will be used to indicate single indeterminates. This convention is however only for indeterminates in the SoS proofs. For other variables, both capital and small letters will be used. Also, we will consider polynomials on the indeterminates occurring in the SoS proofs. Whenever we refer to such polynomials without an explicit reference to the underlying indeterminates, the set of indeterminates will be clear from the context.

5.1 Useful facts in SoS hierarchy

Fact 5.3. *If $A \vdash_d p \geq 0$ and $A \vdash_d q \geq 0$, then $A \vdash_d p+q \geq 0$.*

Fact 5.4. *If $A = \{-1 \leq y \leq 1\}$,*

- *if k is an even integer, $A \vdash_{k+1} 0 \leq y^k \leq 1$,*
- *if k is an odd integer, $A \vdash_k -1 \leq y^k \leq 1$.*

Proof. Note that for $k = 0, 1$, the conclusion is trivially true. For $k = 2$, note that trivially, $y^2 \geq 0$. So, we begin by observing that (from [49])

$$1 - y^2 = \frac{1}{2}(1+y)^2(1-y) + \frac{1}{2}(1+y)(1-y)^2\,,$$

and hence $0 \leq y \leq 1 \vdash_3 y^2 \leq 1$. This finishes the case for $k = 2$. For the remaining cases, we use induction. We first consider the case when $k > 2$ is even. Then, trivially, we have $y^k \geq 0$. Also, observe that $1 - y^k = y^2(1 - y^{k-2}) + (1 - y^2)$. Hence, by induction hypothesis, we have $A \vdash_{k+1} y^k \leq 1$.

Next, consider the case when $k > 2$ is odd. Again, as $1 - y^k = y^2(1 - y^{k-2}) + (1 - y^2)$, hence by induction hypothesis, we get $A \vdash_k y^k \leq 1$. Also, note that $1 + y^k = y^2(1 + y^{k-2}) + (1 - y^2)$. Hence, again, by induction hypothesis, we get $A \vdash_k -1 \leq y^k$ □

Fact 5.5. $-1 \leq y \leq 1 \vdash_5 y^4 \leq y^2$.

Proof.

$$y^2 - y^4 = \frac{1}{2}y^2(1+y)^2(1-y) + \frac{1}{2}(1+y)(1-y)^2y^2\,.$$

As the largest degree appearing is 5, this finishes the proof. □

Fact 5.6. *Let $a \leq y \leq b \vdash_d p(y) \geq 0$. Then, for $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that $\forall i \in [k]$, $\lambda_i \geq 0$ and $\sum_{i=1}^k \lambda_i = 1$, we have that*

$$\left\{\bigcup_{i=1}^k a \leq z_i \leq b\right\} \vdash_d p\Big(\sum_{i=1}^k \lambda_i z_i\Big) \geq 0\,.$$

Proof. In the SoS proof of $p(y) \geq 0$, whenever the term $(b - y)$ appears, we simply substitute it by $\sum_{i=1}^k \lambda_i(b - z_i)$. Likewise, whenever the term $(y - a)$ appears, we substitute it by $\sum_{i=1}^k \lambda_i(a - z_i)$. It is easy to see that this substitution shows that

$$\left\{\bigcup_{i=1}^k a \leq z_i \leq b\right\} \vdash_d p\Big(\sum_{i=1}^k \lambda_i z_i\Big) \geq 0\,.$$ □

Fact 5.7. *For $k \geq 3$, $-1 \leq y \leq 1 \vdash_{2k+1} 0 \leq y^{2k} \leq y^4$.*

Proof. $1 - y^2 = (1/2)(1+y)^2(1-y) + (1/2)(1+y)(1-y)^2$ and hence $-1 \leq y \leq 1 \vdash_3 y^2 \leq 1$. As a consequence, for any $j \geq 1$. we can get that $-1 \leq y \leq 1 \vdash_{2j+1} y^{2j} \leq y^{2j-2}$. Summing all the inequalities as j variables from $j = 3$ to $j = k$, we get the stated inequalities. □

Fact 5.8. *For integers $m,n \geq 2$,*

$$\{-1 \leq y \leq 1, -1 \leq z \leq 1\} \vdash_{1+\max\{2m,2n\}} -(y^4+z^4) \leq y^m z^n \leq (y^4+z^4).$$

Proof. $\vdash_{\max\{2m,2n\}} y^{2m}+z^{2n} \geq y^m z^n$. Also, using Fact 5.7, $-1 \leq y \leq 1 \vdash_{2m+1} y^4 \geq y^{2m}$. Similarly, we have $-1 \leq z \leq 1 \vdash_{2n+1} z^4 \geq z^{2n}$. Combining these, we have

$$\{-1 \leq y \leq 1, -1 \leq z \leq 1\} \vdash_{1+\max\{2m,2n\}} y^m z^n \leq (y^4+z^4).$$

Replacing y by $-y$ and z by $-z$, we can similarly get

$$\{-1 \leq y \leq 1, -1 \leq z \leq 1\} \vdash_{1+\max\{2m,2n\}} -y^m z^n \leq (y^4+z^4).$$

This completes the proof. □

Fact 5.9. *For any odd integer $n \geq 3$,*

$$\{-1 \leq y \leq 1, -1 \leq z \leq 1\} \vdash_{n+2} -(y^4+z^4) \leq yz^n \leq (y^4+z^4).$$

Proof. We use A to denote $\{-1 \leq y \leq 1, -1 \leq z \leq 1\}$. We begin with the case $n=3$. We first use Fact 3.10 from [49] which states that $A \vdash_4 yz^3 \leq y^4+z^4$. This replaces one side of the inequality. We can replace y by $-y$ to get the other inequality. This proves the case $n=3$. For $n>3$, we begin by observing that the case $n=1$ implies $A \vdash_{n+1} yz^n \leq y^4 z^{n-3} + z^{n+1}$. Now, using Fact 5.4 (Item 1), we get $A \vdash_{n+2} z^{n+1} \leq z^4$. And similarly, we get $A \vdash_{n-2} z^{n-3} \leq 1$ and hence $A \vdash_{n+2} z^{n-3} y^4 \leq y^4$. Combining these, we get that $A \vdash_{n+2} yz^n \leq y^4+z^4$. Replacing y by $-y$, we get the other side. □

Fact 5.10. *Let $A \vdash_{d_1} 0 \leq x \leq 1$ and $A \vdash_{d_2} -z \leq y \leq z$ where $z \in \mathbb{M}_{\leq d_3}[X]$. Then,*

$$A \vdash_{d_1+\max\{d_2,d_3\}} -z \leq xy \leq z.$$

Proof. Note that $z - xy = z(1-x) + x(z-y)$. Now, $A \vdash_{d_1+d_2} x(z-y) \geq 0$ and $A \vdash_{d_3+d_1} z(1-x) \geq 0$. Combining these, we get that $A \vdash_{d_1+\max\{d_2,d_3\}} xy \leq z$. Flipping y to $-y$, we get the other inequality. □

Moving on from elementary inequalities, we have a bounded-degree hypercontractive inequality due to Barak *et al*. This inequality will be used in place of the Bonami-Beckner hypercontractive inequality (Theorem 4.6) in the SoS version of the error analysis.

Fact 5.11. [3] *Let $n,d \in \mathbb{N}$ and $d \leq n$. For every $S \subseteq [n]$ such that $|S| \leq d$, we have an indeterminate $\widehat{\ell}(S)$. For $x \in \{-1,1\}^n$, define*

$$\ell(x) = \sum_{S \subseteq [n]:|S| \leq d} \widehat{\ell}(S) \chi_S(x).$$

Then,

$$\vdash_4 \mathop{\mathbf{E}}_{x \in \{-1,1\}^n} [\ell^4(x)] \leq 9^d \left(\mathop{\mathbf{E}}_{x \in \{-1,1\}^n} [\ell^2(x)] \right)^2.$$

We next state the following important theorem of Putinar [25] which shall be repeatedly used in the SoS proof of "Majority is Stablest."

Theorem 5.12. [25] *Let $A = \{p_1(X) \geq 0, \ldots, p_m(X) \geq 0\}$ and define $\mathbb{M}(A)$ to be the set of all polynomials expressible as $\sum_{i=1}^n r_i p_i + r_0$ where $r_0, \ldots, r_m \in \mathbb{M}[X]$. Assume that $\exists q \in \mathbb{M}(A)$ such that the set $\{X : q(X) \geq 0\}$ is compact. If $p > 0$ on the set $\mathbb{V}(A)$, then $p \in \mathbb{M}(A)$.*

In the SoS proof terminology, Putinar's theorem has the following consequence. For the convenience of the reader, we recall that $\mathbb{V}(A)$ is the set of x where the constraints defined by A hold.

Corollary 5.13. *Let $X = (x_1, x_2)$ and $A = \{x_1 \geq \varepsilon, x_2 \geq \varepsilon, x_1 \leq 1-\varepsilon, x_2 \leq 1-\varepsilon\}$. Then, for any $p(X)$ such that $p(X) \geq \varepsilon$ for all $X \in \mathbb{V}(A)$, there exists an integer $d = d(p)$ such that $A \vdash_d p \geq \varepsilon/2$.*

Proof. We can define the polynomials $p_1 = x_1 - \varepsilon$, $p_2 = 1 - \varepsilon - x_1$, $p_3 = x_2 - \varepsilon$ and $p_4 = 1 - \varepsilon - x_2$. Now, note that $q(x,y)$ defined as

$$\begin{aligned} q(x_1, x_2) &= (1-\varepsilon-x_1)^2 \cdot p_1 + (x_1-\varepsilon)^2 \cdot p_2 + (1-\varepsilon-x_2)^2 \cdot p_3 + (x_2-\varepsilon)^2 \cdot p_4 \\ &= (1-2\varepsilon)\left(-\left(x_1 - \frac{1}{2}\right)^2 - \left(x_2 - \frac{1}{2}\right)^2 + \frac{1}{4} - 2\varepsilon(1-\varepsilon)\right). \end{aligned}$$

Clearly, $q(x_1, x_2) \in \mathbb{M}(A)$ and that the set $\{(x_1, x_2) : q(x_1, x_2) \geq 0\}$ is a compact set. As a consequence, we can apply Theorem 5.12 to get that there is an integer $d = d(p)$ such that $A \vdash_d p - \varepsilon/2 \geq 0$. This implies $A \vdash p \geq \varepsilon/2$. $\square$

As a key step in one of our proofs, we will also require a matrix version of Putinar's Positivstellensatz (see [37] for details). A matrix $\Gamma \in (\mathbb{R}[X])^{p \times p}$ is said to be a sum-of-squares if there exists $B \in (\mathbb{R}[X])^{p \times q}$ (for some $q \in \mathbb{N}$) such that $B \cdot B^T = \Gamma$.

Theorem 5.14. *Let $A = \{p_1(X) \geq 0, \ldots, p_m(X) \geq 0\}$ be satisfying the conditions in the hypothesis of Theorem 5.12. Let $\Gamma \in (\mathbb{R}[X])^{p \times p}$ be a symmetric matrix and $\delta > 0$ be such that $\Gamma \succeq \delta I$ on the set $\mathbb{V}(A)$. Then, $\Gamma = \Gamma_0(X) + \sum_{i=1}^m \Gamma_i(X) \cdot p_i(X)$ where $\Gamma_0, \ldots, \Gamma_m$ are sum-of-squares.*

In the same way that Corollary 5.13 followed from Theorem 5.12, we have the following consequence of Theorem 5.14.

Corollary 5.15. *Let $X = (x_1, \ldots, x_n)$ and $A = \{p_1(X) \geq 0, \ldots, p_m(X) \geq 0\}$ be satisfying the conditions in Theorem 5.12. Let $\Gamma \in (\mathbb{R}[X])^{p \times p}$ be such that for $x \in \mathbb{V}(A)$, $\Gamma \succeq \delta I$ for some $\delta > 0$. Let $v \in (\mathbb{R}[X])^p$. Then, if $p = v^T \cdot \Gamma \cdot v$, then $p \in \mathbb{M}(A)$.*

Proof. First, by applying Theorem 5.14, we get $\Gamma = \Gamma_0(X) + \sum_{i=1}^m \Gamma_i(X) \cdot p_i(X)$. Let us assume that $\Gamma_i = B_i^T \cdot B_i$. Then,

$$p = v^T \cdot \Gamma \cdot v = v^T (\Gamma_0(X) + \sum_{i=1}^m \Gamma_i(X) \cdot p_i(X)) v = (B_0 \cdot v)^T \cdot (B_0 \cdot v) + \sum_{i=1}^m (B_0 \cdot v)^T \cdot (B_0 \cdot v) \cdot p_i(x).$$

This proves the claim. $\square$

6 SoS proof of "Majority is Stablest"

The principal theorem of this section is the SoS version of "Majority is Stablest" theorem of [45]. Before we state the theorem, we will need a few definitions. We will consider the indeterminates $f(x)$ (for $x \in \{-1,1\}^n$). The constraints on these indeterminates is given by

$$A_p = \{0 \le f(x) \le 1 : \text{ for all } x \in \{-1,1\}^n\}.$$

As is the case with the usual setting, its helpful to define the Fourier coefficients of f.

$$\text{For } S \subseteq [n], \quad \widehat{f}(S) = \mathop{\mathbf{E}}_{x\in\{-1,1\}^n} f(x)\cdot\chi_S(x) \quad \text{and hence} \quad f(x) = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S(x).$$

Note that $\widehat{f}(S)$ are nothing but linear forms in terms of the original indeterminates. It is also helpful to recall the notion of influences and low-degree influences in this context:

$$\mathrm{Inf}_i(f) = \sum_{S\ni i} \widehat{f}^2(S), \qquad \mathrm{Inf}_i^{\le d}(f) = \sum_{S\ni i:|S|\le d} \widehat{f}^2(S).$$

With this, we state the main theorem of this section.

Theorem 6.1. *For any $\kappa > 0$ and $\rho \in (-1,0)$, $\exists d_0 = d_0(\kappa,\rho)$, $d_1 = d_1(\kappa,\rho)$ and $c = c(\kappa,\rho)$ such that*

$$A_p \vdash_{d_0} \mathop{\mathbf{E}}_{\substack{x\in\{-1,1\}^n \\ y\sim_\rho x}} [f(x)\cdot f(y) + (1-f(x))\cdot(1-f(y))] \ge 1 - \frac{1}{\pi}\arccos\rho - \kappa - c\cdot\left(\sum_{i=1}^n (\mathrm{Inf}_i^{\le d_1} f)^2\right).$$

This is easily seen to be equivalent to the statement of the "Majority is Stablest" theorem of [45]. Before we delve further into the SoS proofs, we will familiarize ourselves with the Fourier machinery in the SoS world. The upshot of the ensuing discussion is that the basic Fourier identities and operations hold without any changes in the SoS world.

We start by verifying that Parseval's identity holds, i. e., for $\{f(x)\}$ and $\{\widehat{f}(S)\}$ defined as above

$$\mathbf{E}[f^2(x)] = \sum_{S\subseteq[n]} \widehat{f}^2(S).$$

Similarly, we can define the noise operator T_ρ here as follows: Given the sequence of indeterminates $\{f(x)\}_{x\in\{-1,1\}^n}$, we define the sequence of indeterminates $\{g(x)\}_{x\in\{-1,1\}^n}$ as

$$g(x) = \mathbf{E}_{y\sim_\rho x}[f(x)]$$

and for every x, use $T_\rho f(x)$ to refer to $g(x)$. It is also easy to check that if we define $\widehat{g}(S) = \mathbf{E}_x[g(x)\cdot\chi_S(x)]$, then $\widehat{g}(S) = \rho^{|S|}\widehat{f}(S)$.

6.1 Smoothening the function

For our purposes, it is necessary to do a certain smoothening of the function f. In particular, we start by considering a new function f_1, i. e., we create a new sequence of indeterminates defined by $f_1(x) = (1-\varepsilon)f(x) + \varepsilon/2$ for some $\varepsilon > 0$. The value of ε shall be fixed later. We observe that

$$\begin{aligned} A_p &\vdash_1 \cup_{x \in \{-1,1\}^n} \{\varepsilon \le f_1(x) \le 1-\varepsilon\}, \\ \widehat{f_1}(S) &= (1-\varepsilon)\hat{f}(S) + (\varepsilon/2) \cdot \mathbf{1}_{S=\Phi}. \end{aligned} \tag{6.1}$$

We begin by making the following claim.

Claim 6.2.

$$A_p \vdash_2 f(x)f(y) - 2\varepsilon \le f_1(x)f_1(y) \le f(x)f(y) + 2\varepsilon.$$

Proof. Begin by noting that $1 - f(x)f(y) = (1-f(x))f(y) + (1-f(y))$ and hence $A_p \vdash_2 0 \le f(x)f(y) \le 1$. Also,

$$f_1(x)f_1(y) - f(x)f(y) = (\varepsilon^2 - 2\varepsilon)f(x)f(y) + \frac{\varepsilon^2}{4} + \varepsilon(1-\varepsilon)(f(x)+f(y)).$$

Using, $A_p \vdash_2 0 \le f(x)f(y)$, we get that

$$A_p \vdash_2 f_1(x)f_1(y) - f(x)f(y) \le 2(1-\varepsilon)\varepsilon + \frac{\varepsilon^2}{4} \le 2\varepsilon.$$

On the other hand, using $A_p \vdash_2 f(x)f(y) \le 1$, we get that

$$A_p \vdash_2 f(x)f(y) - f_1(x)f_1(y) \le 2\varepsilon - \varepsilon^2 \le 2\varepsilon. \qquad \square$$

The next stage of smoothening is done by defining $g = T_{1-\eta} f_1$ for some $\eta > 0$. Again, the value of η will be fixed later. We observe that the following relations hold:

$$\begin{aligned} \bigcup_{x \in \{-1,1\}^n} \{\varepsilon \le f_1(x) \le 1-\varepsilon\} &\vdash_1 \bigcup_{x \in \{-1,1\}^n} \{\varepsilon \le g(x) \le 1-\varepsilon\}, \\ \widehat{g}(S) &= (1-\varepsilon)(1-\eta)^{|S|}\hat{f}(S) + (\varepsilon/2) \cdot \mathbf{1}_{S=\Phi}. \end{aligned} \tag{6.2}$$

Also, observe that

$$\mathbf{E}_{x,y \sim_\rho x}[f_1(x) \cdot f_1(y)] = \mathbf{E}_{x,y \sim_{\rho'} x}[g(x) \cdot g(y)]$$

where $\rho' = \rho/(1-\eta)^2$. Of course, this imposes the condition $|\rho| \le |1-\eta|^2$. So, we have to choose η to be small enough compared to $|\rho|$. Now, define the constraint set

$$A'_p = \bigcup_{x \in \{-1,1\}^n} \{\varepsilon \le g(x) \le 1-\varepsilon\}.$$

So, we summarize the discussion of this subsection in the following two claims.

Claim 6.3. *For any q and $d \in \mathbb{N}$, if $A'_p \vdash_d q \ge 0$, then $A_p \vdash_d q \ge 0$.*

The proof of the above is obtained by combining (6.1) and (6.2) with the third bullet of Fact 5.2. Similarly, using Claim 6.2 and

$$\mathbf{E}_{x,y\sim_\rho x}[f_1(x)\cdot f_1(y)] = \mathbf{E}_{x,y\sim_{\rho'} x}[g(x)\cdot g(y)]\,,$$

we get the next claim:

Claim 6.4.

$$A_p \vdash_2 \mathbf{E}_{x,y\sim_\rho x}[f(x)\cdot f(y)] \geq \mathbf{E}_{x,y\sim_{\rho'} x}[g(x)\cdot g(y)] - 2\varepsilon\,.$$

Thus, the above two claims mean that from now on, we will work with A'_p and aim to prove a lower bound on

$$\mathbf{E}_{x,y\sim_{\rho'} x}[g(x)\cdot g(y)]\,.$$

At this stage, let $\tilde{J}_{\rho'}$ be the approximation obtained from Claim B.1 with parameter $\varepsilon > 0$ and $\delta = \varepsilon$. For the sake of brevity, we indicate this by $\tilde{J}$ itself. The following claim allows us to compare the terms $x\cdot y$ and $\tilde{J}(x,y)$.

Claim 6.5. *For any $\varepsilon > 0$, such that $\rho' \in (-1,0)$ and $\tilde{J}$ is as described above, there is a $d_\alpha = d_\alpha(\varepsilon,\rho')$ such that,*

$$\{\varepsilon \leq x \leq 1-\varepsilon, \varepsilon \leq y \leq 1-\varepsilon\} \vdash_{d_\alpha} x\cdot y \geq \tilde{J}(x,y) - 2\varepsilon.$$

Proof. Note that for $(x,y)\in(0,1)^2$, $J_0(x,y) = xy$ and hence by Slepian's lemma, we get that if $\rho' < 0$, then $xy \geq J_{\rho'}(x,y)$. Now, by definition, we have that for $(x,y)\in[\varepsilon,1-\varepsilon]^2$, $xy \geq \tilde{J}(x,y)-\varepsilon$. In other words, if we define the polynomial $p(x,y) = xy - \tilde{J}(x,y) + 2\varepsilon$, then we know that for $(x,y)\in[\varepsilon,1-\varepsilon]^2$, $p(x,y)\geq\varepsilon$. We can thus apply Corollary 5.13 to get that there is an integer $d_\alpha = d_\alpha(\varepsilon,\rho')$ such that for $(x,y)\in[\varepsilon,1-\varepsilon]^2$, for $\rho'\in(0,1)$,

$$\{\varepsilon \leq x \leq 1-\varepsilon, \varepsilon \leq y \leq 1-\varepsilon\} \vdash_{d_\alpha} p \geq 0\,.$$

Expanding p, finishes the proof. □

6.2 Taylor's theorem in the SoS world

Following the proof of "Majority is Stablest," we now need to prove a Taylor's theorem in the SoS hierarchy. The following lemma is the SoS analogue of Claim 2.5.

Lemma 6.6. *Define a sequence of indeterminates $\{h_0(1),h_0(-1),h_1(1),h_1(-1)\}$. Let A be a set of constraints defined as $A = \bigcup_{i,j\in\{0,1\}}\{\varepsilon \leq h_i(j) \leq 1-\varepsilon\}$. For any $\varepsilon > 0$, $\rho'\in(-1,0)$, $\exists c_\gamma = c_\gamma(\varepsilon,\rho')$ and $\exists d_\gamma = d_\gamma(\varepsilon,\rho')$ such that*

$$A \vdash_{d_\gamma} \mathop{\mathbb{E}}_{\substack{x\in_R\{-1,1\}\\ y\sim_{\rho'}x}}[\widetilde{J}(h_0(x),h_1(y))] \geq \widetilde{J}(\widehat{h_0}(0),\widehat{h_1}(0)) - \varepsilon\cdot(\widehat{h_0}^2(1)+\widehat{h_1}^2(1)) - c_\gamma\cdot(\widehat{h_0}^4(1)+\widehat{h_1}^4(1))\,,$$

where

$$\widehat{h_i}(j) = \frac{h_i(1)+(-1)^j\cdot h_i(-1)}{2}$$

for $i,j\in\{0,1\}$. Further, $c_\gamma(\varepsilon,\rho')$ is a continuous function of ε and ρ'.

Note that while the error term in Claim 2.5 was third-order, the error term in Lemma 6.6 consists of a small second-order term and a large fourth-order term. This sort of error term (which was used also in [49]) is more convenient in the SoS world, because it avoids the need to consider absolute values.

The proof of Lemma 6.6 is fairly long, since it must be written in the restrictive SoS proof system. To give some motivation for the proof, let us first give a short outline. First of all, if we restrict to the first- and second-order terms of $\widetilde{J}$, then the claimed inequality is true even without the fourth-order error term. By the completeness of the SoS proof system (Corollary 5.15), there must therefore be a constant-degree SoS proof of this fact. Most of the proof of Lemma 6.6 goes into showing that the higher order terms of $\widetilde{J}$ can be bounded by a fourth-order error term. This is of course trivial to show in the usual proof system, but takes some work to write as an SoS proof.

Proof of Lemma 6.6. We start by noting that since $\widetilde{J}$ is a symmetric polynomial, hence we can write

$$\widetilde{J}(x,y) = \sum_{m,n:m+n\le K} \mu_{m,n} x^m y^n .$$

Here, we assume that K is the degree of $\widetilde{J}$ and c is the maximum absolute value of any coefficient. We next make the following claim.

Claim 6.7.

$$\mathop{\mathbb{E}}_{\substack{x\in_R\{-1,1\}\\ y\sim_{\rho'}x}}[\widetilde{J}(h_0(x),h_1(y))] = \sum_{m,n:m+n \text{ is even}} \nu_{m,n}\cdot \widehat{h_0}^m(1)\cdot \widehat{h_1}^n(1)\cdot\left(\frac{1+\rho'}{2}+(-1)^m\cdot\frac{1-\rho'}{2}\right),$$

where

$$\nu_{m,n} = \sum_{m_1\ge m;n_1\ge n} \mu_{m_1,n_1}\cdot \widehat{h_0}^{m_1-m}(0)\cdot \widehat{h_1}^{n_1-n}(0)\cdot \binom{m_1}{m}\binom{n_1}{n}.$$

Proof. We begin by observing that

$$\widetilde{J}(\widehat{h_0}(0)+x,\widehat{h_1}(0)+y) = \sum_{m,n}\mu_{m,n}\cdot(\widehat{h_0}(0)+x)^m\cdot(\widehat{h_1}(0)+y)^n = \sum_{m,n}\nu_{m,n}\cdot x^m\cdot y^n .$$

As a consequence, we get that

$$\widetilde{J}(\widehat{h_0}(0)+\widehat{h_0}(1),\widehat{h_1}(0)+\widehat{h_1}(1)) = \sum_{m,n}\nu_{m,n}\cdot\widehat{h_0}^m(1)\cdot\widehat{h_1}^n(1),$$
$$\widetilde{J}(\widehat{h_0}(0)-\widehat{h_0}(1),\widehat{h_1}(0)-\widehat{h_1}(1)) = \sum_{m,n}(-1)^{m+n}\nu_{m,n}\cdot\widehat{h_0}^m(1)\cdot\widehat{h_1}^n(1).$$

Adding these equations, we get

$$\widetilde{J}(\widehat{h_0}(0)+\widehat{h_0}(1),\widehat{h_1}(0)+\widehat{h_1}(1))+\widetilde{J}(\widehat{h_0}(0)-\widehat{h_0}(1),\widehat{h_1}(0)-\widehat{h_1}(1)) = 2\cdot\sum_{\substack{m,n:\\ m+n \text{ is even}}}\nu_{m,n}\cdot\widehat{h_0}^m(1)\cdot\widehat{h_1}^n(1). \quad (6.3)$$

Similarly, we have that

$$\widetilde{J}(\widehat{h_0}(0)-\widehat{h_0}(1),\widehat{h_1}(0)+\widehat{h_1}(1)) = \sum_{m,n}(-1)^m \nu_{m,n}\cdot \widehat{h_0}^m(1)\cdot \widehat{h_1}^n(1)\,,$$
$$\widetilde{J}(\widehat{h_0}(0)+\widehat{h_0}(1),\widehat{h_1}(0)-\widehat{h_1}(1)) = \sum_{m,n}(-1)^n \nu_{m,n}\cdot \widehat{h_0}^m(1)\cdot \widehat{h_1}^n(1)\,.$$

Thus,

$$\widetilde{J}(\widehat{h_0}(0)-\widehat{h_0}(1),\widehat{h_1}(0)+\widehat{h_1}(1))+\widetilde{J}(\widehat{h_0}(0)+\widehat{h_0}(1),\widehat{h_1}(0)-\widehat{h_1}(1)) = 2\cdot \sum_{\substack{m,n:\\ m+n \text{ is even}}}(-1)^m \nu_{m,n}\cdot \widehat{h_0}^m(1)\cdot \widehat{h_1}^n(1)\,. \tag{6.4}$$

Hence, combining (6.3) and (6.4), we get that

$$\mathop{\mathbf{E}}_{\substack{x\in_R\{-1,1\}\\ y\sim_{\rho'}x}}[\widetilde{J}(h_0(x),h_1(y))] = \sum_{\substack{m,n:\\ m+n \text{ is even}}} \nu_{m,n}\cdot \widehat{h_0}^m(1)\cdot \widehat{h_1}^n(1)\cdot\left(\frac{1+\rho'}{2}+(-1)^m\cdot\frac{1-\rho'}{2}\right). \qquad \square$$

Next, we note that $\nu_{0,0}=\widetilde{J}(\widehat{h_0}(0),\widehat{h_1}(0))$. Thus, we get that

$$\mathop{\mathbb{E}}_{\substack{x\in_R\{-1,1\}\\ y\sim_{\rho'}x}}[\widetilde{J}(h_0(x),h_1(y))]-\widetilde{J}(\widehat{h_0}(0),\widehat{h_1}(0)) = \sum_{\substack{m,n:m+n \text{ is even}\\ K\geq m+n\geq 2}} \nu_{m,n}\cdot \widehat{h_0}^m(1)\cdot \widehat{h_1}^n(1)\cdot\left(\frac{1+\rho'}{2}+(-1)^m\cdot\frac{1-\rho'}{2}\right). \tag{6.5}$$

We first make the following claim which bounds the terms when $m+n\geq 4$.

Claim 6.8.

$$A\vdash_{2K+3} Y \geq \sum_{\substack{m,n:m+n \text{ is even}\\ \text{and } m+n\geq 4}} \nu_{m,n}\cdot \widehat{h_0}^m(1)\cdot \widehat{h_1}^n(1)\cdot\left(\frac{1+\rho'}{2}+(-1)^m\cdot\frac{1-\rho'}{2}\right)\geq -Y\,,$$

where $Y=2cK^4 2^{2K}(\widehat{h_0}^4(1)+\widehat{h_1}^4(1))$.

Proof. For $m+n\geq 4$ and $m_1\geq m$ and $n_1\geq n$, we define $\Gamma_{m_1,n_1,m,n}$ as follows:

$$\Gamma_{m_1,n_1,m,n} = \mu_{m_1,n_1}\binom{m_1}{m}\binom{n_1}{n}\cdot \widehat{h_0}^m(1)\cdot \widehat{h_1}^n(1)\cdot \widehat{h_0}^{m_1-m}(0)\cdot \widehat{h_1}^{n_1-n}(0)\cdot\left(\frac{1+\rho'}{2}+(-1)^m\cdot\frac{1-\rho'}{2}\right).$$

Next, define the set of constraints A_m as

$$A_m=\{0\leq \widehat{h_0}(0)\leq 1\,,\ 0\leq \widehat{h_1}(0)\leq 1\,,\ -1\leq \widehat{h_0}(1)\leq 1\,,\ -1\leq \widehat{h_1}(1)\leq 1\}\,.$$

Now, it is easy to see that $A\vdash_1 A_m$. Hence, by using the third bullet of Fact 5.2, if for any p and $d\in\mathbb{N}$, $A_m\vdash_d p\geq 0$, then $A\vdash_d p\geq 0$. We shall be using this fact throughout this proof. Applying Fact 5.4, we get that

$$A\vdash_{m_1-m+1} 0\leq \widehat{h_0}^{m_1-m}(0)\leq 1\,,\qquad A\vdash_{n_1-n+1} 0\leq \widehat{h_1}^{n_1-n}(0)\leq 1\,,$$

and hence we have that

$$A \vdash_{m_1+n_1-m-n+2} \quad 0 \le \widehat{h_0}^{m_1-m}(0) \cdot \widehat{h_1}^{n_1-n}(0) \le 1\,. \tag{6.6}$$

Now, we consider two possibilities: Either $m = n = 2$ or $\max\{m,n\} \ge 3$. In the first case, using Fact 5.8, we get that

$$A \vdash_5 -(\widehat{h_0}^4(1) + \widehat{h_1}^4(1)) \le \widehat{h_0}^2(1) \cdot \widehat{h_1}^2(1) \le \widehat{h_0}^4(1) + \widehat{h_1}^4(1)\,. \tag{6.7}$$

Next, consider the other case, i. e., when $\max\{m,n\} \ge 3$. Without loss of generality, assume $m \ge 3$. Then, by Fact 5.9, we get that

$$A \vdash_{m+2} -(\widehat{h_0}^4(1) + \widehat{h_1}^4(1)) \le \widehat{h_0}^m(1) \cdot \widehat{h_1}(1) \le \widehat{h_0}^4(1) + \widehat{h_1}^4(1)\,.$$

Next, we use Fact 5.10 to get that

$$A \vdash_{m+n+1} -(\widehat{h_0}^4(1) + \widehat{h_1}^4(1)) \le \widehat{h_0}^m(1) \cdot \widehat{h_1}^n(1) \le \widehat{h_0}^4(1) + \widehat{h_1}^4(1)\,. \tag{6.8}$$

Thus, combining (6.7) and (6.8), we get that for any m and n such that $m+n \ge 4$,

$$A \vdash_{m+n+1} -(\widehat{h_0}^4(1) + \widehat{h_1}^4(1)) \le \widehat{h_0}^m(1) \cdot \widehat{h_1}^n(1) \le \widehat{h_0}^4(1) + \widehat{h_1}^4(1)\,.$$

Now, combining (6.6) along with an application of Fact 5.10, we get that

$$A \vdash_{m_1+n_1+3} -(\widehat{h_0}^4(1) + \widehat{h_1}^4(1)) \le \widehat{h_0}^m(1) \cdot \widehat{h_1}^n(1) \cdot \widehat{h_0}^{m_1-m}(0) \cdot \widehat{h_1}^{n_1-n}(0) \le \widehat{h_0}^4(1) + \widehat{h_1}^4(1)\,.$$

Now, recalling that $0 \le m_1, n_1, m, n \le K$, $\binom{m_1}{m}, \binom{n_1}{n} \le 2^K$, $|\mu_{m,n}| \le c$ and $|\rho'| \le 1$, we get that

$$A \vdash_{2K+3} -c2^{2K}(\widehat{h_0}^4(1) + \widehat{h_1}^4(1)) \le \Gamma_{m_1,n_1,m,n} \le c2^{2K}(\widehat{h_0}^4(1) + \widehat{h_1}^4(1))\,.$$

As

$$\sum_{\substack{m,n: m+n \text{ is even} \\ \text{and } m+n \ge 4}} \nu_{m,n} \widehat{h_0}^m(1) \widehat{h_1}^n(1) \cdot \left(\frac{1+\rho'}{2} + (-1)^m \cdot \frac{1-\rho'}{2} \right) = \sum_{\substack{m,n: m+n \text{ is even} \\ m+n \ge 4 \\ K \ge m_1 \ge m\ K \ge n_1 \ge n}} \Gamma_{m_1,n_1,m,n}\,,$$

we can conclude that

$$A \vdash_{2K+3} -c2^{2K}K^4 \cdot (\widehat{h_0}^4(1) + \widehat{h_1}^4(1)) \le \sum_{\substack{m,n: m+n \text{ is even} \\ \text{and } m+n \ge 4}} \nu_{m,n} \widehat{h_0}^m(1) \cdot \widehat{h_1}^n(1) \le c2^{2K}K^4 \cdot (\widehat{h_0}^4(1) + \widehat{h_1}^4(1))\,.$$

This completes the proof of Claim 6.8. □

Finally, we need to consider the terms when $m+n=2$. This is where we will use Putinar's theorem. Note that

$$\sum_{m+n=2} \nu_{m,n} \cdot \widehat{h_0}^m(1) \cdot \widehat{h_1}^n(1) \cdot \left(\frac{1+\rho'}{2} + (-1)^m \cdot \frac{1-\rho'}{2} \right) = \nu_{2,0} \cdot \widehat{h_0}^2(1) + \nu_{0,2} \cdot \widehat{h_1}^2(1) + \rho' \nu_{1,1} \cdot \widehat{h_0}(1) \cdot \widehat{h_1}(1)\,.$$

For the sake of brevity, call the above quantity Λ. Next, we observe that at $x = \widehat{h_0}(0), y = \widehat{h_1}(0)$,

$$\frac{\partial^2 \widetilde{J}(x,y)}{\partial x^2} = 2\nu_{2,0}\,, \quad \frac{\partial^2 \widetilde{J}(x,y)}{\partial y^2} = 2\nu_{0,2}\,, \quad \frac{\partial^2 \widetilde{J}(x,y)}{\partial x \partial y} = \nu_{1,1}\,.$$

To see this, note that

$$\left.\frac{\partial^2 \widetilde{J}(x,y)}{\partial x^2}\right|_{x=\widehat{h_0}(0),\ y=\widehat{h_1}(0)} = \left.\frac{\partial^2 \widetilde{J}(\widehat{h_0}(0)+x', \widehat{h_1}(0)+y')}{\partial x'^2}\right|_{x'=0,\ y'=0}.$$

However, the quantity on the right side is simply twice the coefficient of x'^2 in the polynomial $\widetilde{J}(\widehat{h_0}(0)+x', \widehat{h_1}(0)+y')$ which is exactly $2\nu_{2,0}$. The other equalities follow similarly. Thus, we get that

$$\Lambda = \frac{1}{2}\left(\frac{\partial^2 \tilde{J}(x,y)}{\partial x^2}\widehat{h_0}^2(1) + \frac{\partial^2 \tilde{J}(x,y)}{\partial y^2}\widehat{h_1}^2(1) + 2\rho' \frac{\partial^2 \tilde{J}(x,y)}{\partial x \partial y}\widehat{h_0}(1)\cdot \widehat{h_1}(1)\right).$$

We mention here that in the above equation, the derivatives are evaluated at $x = \widehat{h_0}(0),\ y = \widehat{h_1}(0)$. We do not specify this in the equation for the sake of compactness. We now make the following claim which gives a lower bound on Λ.

Claim 6.9. *For every $\varepsilon > 0$ and $\rho' \in (-1,0)$, there exists $d'_\gamma = d'_\gamma(\varepsilon, \rho')$ such that*

$$A \vdash_{d'_\gamma} \Lambda \geq -\varepsilon \cdot (\widehat{h_0}^2(1) + \widehat{h_1}^2(1))\,.$$

Proof. We begin by defining the matrix $\widetilde{M}$ as follows:

$$\widetilde{M} = \begin{pmatrix} \frac{\partial^2 \tilde{J}(x,y)}{\partial x^2} & \rho' \frac{\partial^2 \tilde{J}(x,y)}{\partial x \partial y} \\ \rho' \frac{\partial^2 \tilde{J}(x,y)}{\partial x \partial y} & \frac{\partial^2 \tilde{J}(x,y)}{\partial y^2} \end{pmatrix}.$$

Put $\beta = 2\varepsilon$. Now, let us define $\widetilde{M_1} = \widetilde{M} + \beta I$. Using Claim 2.6 and Claim B.1, for $\rho' \in (-1,0)$, and $(x,y) \in [\varepsilon, 1-\varepsilon]^2$, we can say that $\widetilde{M_1} \succeq \varepsilon \cdot I$. Hence, using Corollary 5.15, $\exists d'_\gamma$ such that we have the following

$$\begin{aligned} & A \vdash_{d'_\gamma} \widehat{h_0}^2(1)\left(\frac{\partial^2 \tilde{J}(x,y)}{\partial x^2} + \beta\right) + 2\rho' \cdot \widehat{h_0}(1)\cdot \widehat{h_1}(1) \cdot \frac{\partial^2 \tilde{J}(x,y)}{\partial x \partial y} + \widehat{h_1}^2(1)\left(\frac{\partial^2 \tilde{J}(x,y)}{\partial y^2} + \beta\right) \geq 0 \\ \equiv\ & A \vdash_{d'_\gamma} \widehat{h_0}^2(1)\cdot \frac{\partial^2 \tilde{J}(x,y)}{\partial x^2} + 2\rho' \cdot \widehat{h_0}(1)\cdot \widehat{h_1}(1) \cdot \frac{\partial^2 \tilde{J}(x,y)}{\partial x \partial y} + \widehat{h_1}^2(1)\cdot \frac{\partial^2 \tilde{J}(x,y)}{\partial y^2} \geq -\beta(\widehat{h_0}^2(1) + \widehat{h_1}^2(1))\,. \end{aligned}$$

Dividing by 2 on both sides, finishes the proof. Note that the reason d'_γ depends only on ε and ρ' is because from Corollary 5.15, the degree d'_γ depends on ε, ρ' and the polynomial $\tilde{J}$ and $\tilde{J}$ depends only on ε and ρ'. □

Now, set $d_\gamma = \max\{d'_\gamma, 2K+3\}$. Also, note that both K and c are bounded by continuous functions of ε and ρ' (see Claim B.1), hence c_γ can be chosen to be a continuous function of ρ' and ε and be an upper bound on $2cK^4 2^{2K}$. Combining Claim 6.8 and Claim 6.9 with (6.5), we get Lemma 6.6. □

6.3 Tensorization

We now do a "tensorization" of the inequality in Lemma 6.6 (analogous to Theorem 2.4). Unlike in the base case, this step of the SoS proof follows the non-SoS proof fairly closely. Let $\{\phi(x)\}_{x\in\{-1,1\}^n}$ be a set of indeterminates. We recall that for $y \in \{-1,1\}^i$, we define the set $\{\phi_y(z)\}_{z\in\{-1,1\}^{n-i}}$ of indeterminates as follows: $\phi_y(z) = \phi(z\cdot y)$ where $z\cdot y$ denotes the concatenation of z and y. As before, we can define the Fourier coefficients $\widehat{\phi_y}(S)$ for $S \subseteq [n-i]$ and it is easy to see that they are homogeneous linear forms in the indeterminates $\widehat{\phi}(S)$ (for $S \subseteq [n]$). We now state a few basic properties of the indeterminates $g_y(z)$ and $\widehat{g_y}(S)$. (The proofs are easy and can be filled in by the reader.)

$$A'_p \vdash_1 \bigcup_{i=0}^{n-1} \bigcup_{y\in\{-1,1\}^i} \bigcup_{z\in\{-1,1\}^{n-i}} \{\varepsilon \le g_y(z) \le 1-\varepsilon\}\,, \tag{6.9}$$

$$A'_p \vdash_1 \bigcup_{i=0}^{n-1} \bigcup_{y\in\{-1,1\}^i} \bigcup_{S\subseteq[n-i]} \{-1 \le \widehat{g_y}(S) \le 1\}\,, \tag{6.10}$$

$$\vdash_2 \mathop{\mathbf{E}}_{y\in\{-1,1\}^i}[\widehat{g_y}^2(n-i)] = \sum_{\substack{S\subseteq\{n-i,\dots,n\}\\ n-i\in S}} \widehat{g}^2(S)\,. \tag{6.11}$$

Lemma 6.10. *For the parameters $c_\gamma = c_\gamma(\varepsilon,\rho')$ and $d_\gamma = d_\gamma(\varepsilon,\rho')$ from Lemma 6.6,*

$$A'_p \vdash_{d_\gamma} \mathop{\mathbf{E}}_{\substack{x\in\{-1,1\}^n\\ y\sim_{\rho'}x}}[\widetilde{J}(g(x),g(y))] \ge \widetilde{J}(\mathbf{E}[g(x)],\mathbf{E}[g(y)]) - \varepsilon\left(\sum_{i=0}^{n-1}\mathbf{E}_{z\in\{-1,1\}^i}[\widehat{g}_z^2(n-i)]\right) - c_\gamma\left(\sum_{i=0}^{n-1}\mathbf{E}_{z\in\{-1,1\}^i}[\widehat{g}_z^4(n-i)]\right)\,.$$

Proof. The proof is by induction. For any pair of strings $z_{-1}, z_1 \in \{-1,1\}^i$ and $j \in \{-1,1\}$, we define the indeterminate,

$$h_i(j) = \mathbf{E}_{z\in\{-1,1\}^{n-i-1}}[g(z\cdot j\cdot z_i)]\,.$$

Define

$$A_{z_{-1},z_1} = \bigcup_{j\in\{-1,1\}} \{\varepsilon \le h_i(j) \le 1-\varepsilon\}\,.$$

It is trivial to see that $A'_p \vdash_1 A_{z_{-1},z_1}$. Now, using Lemma 6.6, for any two strings $z_1, z_{-1} \in \{-1,1\}^i$, we get that

$$A_{z_{-1},z_1} \vdash_{d_\gamma} \mathop{\mathbf{E}}_{\substack{x\in\{-1,1\}\\ y\sim_{\rho'}x}} \widetilde{J}(\mathbf{E}[g_{x\cdot z_1}],\mathbf{E}[g_{y\cdot z_{-1}}]) \ge \widetilde{J}(\mathbf{E}[g_{z_1}],\mathbf{E}[g_{z_{-1}}]) - \varepsilon\left(\widehat{g}_{z_1}^2(n-i)+\widehat{g}_{z_{-1}}^2(n-i)\right) - c_\gamma\left(\widehat{g}_{z_{-1}}^4(n-i)+\widehat{g}_{z_1}^4(n-i)\right)\,.$$

As a consequence of the third bullet of Fact 5.2, the left hand side of "$\vdash$" can be replaced by A'_p. Now, for any given $z_1, z_{-1} \in \{-1,1\}^i$, let $d(z_1, z_{-1})$ be its Hamming distance. In particular, if $i = 0$, then $d(z_1, z_{-1}) = 0$. Now, multiply the above inequality by

$$\left(\frac{1+\rho'}{4}\right)^{i-d(z_1,z_{-1})} \cdot \left(\frac{1-\rho'}{4}\right)^{d(z_1,z_{-1})}$$

and add all such inequalities generated by choosing $(z_1, z_{-1}) \in \{-1,1\}^i \times \{-1,1\}^i$ for $0 \le i < n$. It is easy that upon addition, all terms of the form $\widetilde{J}(\mathbf{E}[g_{z_1}], \mathbf{E}[g_{z_{-1}}])$ cancel out except when $z_1, z_{-1} \in \{-1,1\}^n$ or $z_1 = z_{-1} = \phi$. Thus, we get the inequality

$$A'_p \vdash_{d_\gamma} \mathop{\mathbf{E}}_{\substack{x\in\{-1,1\}^n \\ y\sim_{\rho'} x}} [\widetilde{J}(g(x), g(y))] \ge \widetilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) + \text{error terms}$$

where the error terms are the terms coming from $\widehat{g}^2_{z_j}(n-i)$ and $\widehat{g}^4_{z_j}(n-i)$. We now explicitly compute the error terms. First, we sum up the error coming from the term

$$\varepsilon\left(\widehat{g}^2_{z_1}(n-i) + \widehat{g}^2_{z_{-1}}(n-i)\right).$$

For any given $z_1 \in \{-1,1\}^i$, consider the term $\varepsilon \widehat{g}^2_{z_1}(n-i)$. For every $z_{-1} \in \{-1,1\}^i$, it occurs with the factor

$$\left(\frac{1+\rho'}{4}\right)^{i-d(z_1,z_{-1})} \cdot \left(\frac{1-\rho'}{4}\right)^{d(z_1,z_{-1})}.$$

Since there are exactly $\binom{i}{k}$ strings $z_{-1} \in \{-1,1\}^i$ such that $d(z_1, z_{-1}) = k$, hence we get that the total weight associated is

$$\sum_{k=0}^{i} \binom{i}{k} \left(\frac{1+\rho'}{4}\right)^{i-k} \left(\frac{1-\rho'}{4}\right)^{k} = 2^{-i}.$$

Thus, we get that the first kind of error terms contribute

$$\varepsilon\left(\widehat{g}^2_{z_1}(n-i) + \widehat{g}^2_{z_{-1}}(n-i)\right).$$

The calculation of the "fourth degree" error terms is exactly identical resulting in the final theorem. $\square$

We now simplify the error terms. Towards this, note that (6.11) implies that

$$\vdash_2 \varepsilon \cdot \left(\sum_{i=0}^{n-1} \mathbf{E}_{z\in\{-1,1\}^i}[\widehat{g}^2_z(n-i)]\right) = \varepsilon \cdot \left(\sum_{S\neq\phi} \widehat{g}^2(S)\right) \le \varepsilon \cdot \left(\sum_{S} \widehat{g}^2(S)\right) = \varepsilon \cdot \mathop{\mathbf{E}}_{x\in\{-1,1\}^n}[g^2(x)].$$

Further, $A'_p \vdash_3 \mathbf{E}_{x\in\{-1,1\}^n}[g^2(x)] \le 1$ (using Fact 5.4). Thus, we get that

$$A'_p \vdash_{d_\gamma} \mathop{\mathbf{E}}_{\substack{x\in\{-1,1\}^n \\ y\sim_{\rho'} x}} [\widetilde{J}(g(x), g(y))] \ge \widetilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) - \varepsilon - c_\gamma \left(\sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}^4_x(n-i)]\right). \tag{6.12}$$

6.4 Bounding the error terms

Thus, all we are left to bound is the "degree-4" term. We briefly describe why one has to be careful to get a (meaningful) upper bound here. The reason is that the obvious strategy to do this is to break g into high degree and low-degree parts based on the noise parameter (call them h and ℓ). Very naively, this gives an error which is the sum of $\mathbf{E}_x\widehat{h}_x^4(n-i)$ and $\mathbf{E}_x\widehat{\ell}_x^4(n-i)$ (up to constant factors). The term $\mathbf{E}_x\widehat{\ell}_x^4(n-i)$ can be easily bounded using hypercontractivity. However, there does not seem to be obvious way to bound the term $\mathbf{E}_x\widehat{h}_x^4(n-i)$. This is in spite of the fact that $\mathbf{E}_x\widehat{h}_x^2(n-i)$ is small. We now show how to get around this problem.

We define $d_\eta = (1/\eta)\cdot\log(1/\eta)$. Now, define the sequence of indeterminates $\{h(x)\}_{x\in\{-1,1\}^n}$ and $\{\ell(x)\}_{x\in\{-1,1\}^n}$ as follows:

$$h(x) = \sum_{|S|>d_\eta} \widehat{g}(S)\chi_S(x)\,, \qquad \ell(x) = \sum_{|S|\le d_\eta} \widehat{g}(S)\chi_S(x)\,.$$

By the way it is defined, it is clear that $\vdash_1 h(x)+\ell(x) = g(x)$. Now, we can analyze the term

$$\mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^4(n-i)]$$

as

$$\begin{aligned}
\vdash_4 \sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^4(n-i)] &= \sum_{i=0}^{n-1}(\mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^3(n-i)(\widehat{h}_x(n-i)+\widehat{\ell}_x(n-i))]) \\
&= \sum_{i=0}^{n-1}\Big(\mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^3(n-i)\widehat{h}_x(n-i)] + \mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^2(n-i)\widehat{\ell}_x^2(n-i)] \\
&+ \mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^2(n-i)\widehat{\ell}_x(n-i)\widehat{h}_x(n-i)]\Big)\,. \qquad (6.13)
\end{aligned}$$

We begin by stating the following useful fact:

Fact 6.11. $A_p \vdash_3 \sum_{i=0}^{n-1}\mathbf{E}_{x\in\{-1,1\}^i}\widehat{h}_x^2(n-i) \le \eta$.

Proof.

$$\vdash_2 \sum_{i=0}^{n-1}\mathbf{E}_{x\in\{-1,1\}^i}\widehat{h}_x^2(n-i) = \sum_{S\ne\phi}\widehat{h}^2(S) = \sum_{|S|>d_\eta}(1-\eta)^{d_\eta}\widehat{f}^2(S) \le \eta\cdot\Big(\sum_{|S|>d_\eta}\widehat{f}^2(S)\Big) \le \eta\cdot\Big(\sum_{|S|}\widehat{f}^2(S)\Big).$$

Here the first equality uses (6.11) while the second uses the definition of h.

$$A_p \vdash_3 \Big(\sum_{|S|}\widehat{f}^2(S)\Big) = \mathbf{E}_{x\in\{-1,1\}^n}[f^2(x)] \le 1\,.$$

Combining the two facts finishes the proof. $\square$

We now bound the terms appearing in (6.13). This is done in Claims 6.12, 6.13 and 6.14.

Claim 6.12. $A_p \vdash_7 \sum_{i=0}^{n-1}\mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^3(n-i)\widehat{h}_x(n-i)] \le \sqrt{\eta}$.

Proof. For any $\eta > 0$, we have that

$$\vdash_6 \sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^3(n-i)\widehat{h}_x(n-i)] \le \frac{\sqrt{\eta}}{2}\left(\sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}\widehat{g}_x^6(n-i)\right) + \frac{\left(\sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}\widehat{h}_x^2(n-i)\right)}{2\sqrt{\eta}}.$$

Next, recall that using Fact 6.11, we have

$$A_p \vdash_3 \sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}\widehat{h}_x^2(n-i) \le \eta\,.$$

Similarly, from (6.10) and Claim 6.3, we have that $A_p \vdash -1 \le \hat{g}_x(n-i) \le 1$. This in turn implies $A_p \vdash_7 \widehat{g}_x^6(n-i) \le \widehat{g}_x^2(n-i)$ (combining Fact 5.5 and Fact 5.7). Combining all the above, we get

$$A_p \vdash_7 \sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^3(n-i)\widehat{h}_x(n-i)] \le \frac{\sqrt{\eta}}{2}\cdot\left(\sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}\widehat{g}_x^2(n-i)\right) + \frac{\sqrt{\eta}}{2}\,.$$

However, we have that

$$A_p \vdash_3 \sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}\hat{g}_x^2(n-i) = \sum_{|S|>0} \widehat{g}^2(S) \le \mathbf{E}[g^2(x)] \le 1\,.$$

This proves the claim. $\square$

Claim 6.13.

$$A_p \vdash_5 \sum_{i=0}^{n-1} \mathop{\mathbf{E}}_{x\in\{-1,1\}^i}[\widehat{g}_x^2(n-i)\widehat{\ell}_x^2(n-i)] \le \frac{\sqrt{\eta}}{2} + \frac{9^{d_\eta}}{2\sqrt{\eta}}\left(\sum_{i=1}^{n}(\mathrm{Inf}_i^{\le d_\eta}(f))^2\right).$$

Proof. For $\eta > 0$, we have that

$$\vdash_4 \sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^2(n-i)\widehat{\ell}_x^2(n-i)] \le \frac{\sqrt{\eta}}{2}\left(\sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}\widehat{g}_x^4(n-i)\right) + \frac{\left(\sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}\widehat{\ell}_x^4(n-i)\right)}{2\sqrt{\eta}}.$$

Similarly, using Fact 5.11,

$$\vdash_4 \sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}\widehat{\ell}_x^4(n-i) \le 9^{d_\eta}\cdot\left(\sum_{i=0}^{n-1}\left(\mathbf{E}_{x\in\{-1,1\}^i}\widehat{\ell}_x^2(n-i)\right)^2\right).$$

As in the proof of Claim 6.12, we can show

$$A_p \vdash_5 \sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}\widehat{g}_x^4(n-i) \le 1\,.$$

Combining all the above, we get

$$A_p \vdash_5 \sum_{i=0}^{n-1} \mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^2(n-i)\widehat{\ell}_x^2(n-i)] \le \frac{\sqrt{\eta}}{2} + \frac{9^{d_\eta}}{2\sqrt{\eta}}\left(\sum_{i=0}^{n-1}\left(\mathbf{E}_{x\in\{-1,1\}^i}\widehat{\ell}_x^2(n-i)\right)^2\right). \tag{6.14}$$

Again, observe that

$$\begin{aligned}\vdash_2 \mathbf{E}_{x\in\{-1,1\}^i}\widehat{\ell}_x^2(n-i) &= \sum_{S\subseteq\{n-i,\dots,n\}:n-i\in S}\widehat{\ell}^2(S) \le \sum_{S\subseteq[n]:n-i\in S}\widehat{\ell}^2(S)\\ &= \sum_{\substack{S\subseteq[n]:n-i\in S\\ \text{and } |S|\le d_\eta}}\widehat{g}^2(S) \le \sum_{\substack{S\subseteq[n]:n-i\in S\\ \text{and } |S|\le d_\eta}}\widehat{f}^2(S) \le \mathrm{Inf}_{n-i}^{\le d_\eta}(f).\end{aligned}$$

As a consequence of Fact 5.2, we get

$$\vdash_4 \sum_{i=0}^{n-1}\left(\mathbf{E}_{x\in\{-1,1\}^i}\widehat{\ell}_x^2(n-i)\right)^2 \le \sum_{i=0}^{n-1}\left(\mathrm{Inf}_{n-i}^{\le d_\eta}(f)\right)^2.$$

Combining this with (6.14), we get the final result. □

Claim 6.14.

$$A_p \vdash_9 \sum_{i=0}^{n-1}\mathbf{E}_{x\in\{-1,1\}^i}[\widehat{g}_x^2(n-i)\widehat{h}_x(n-i)\widehat{\ell}_x(n-i)] \le \sqrt{\eta} + \frac{9^{d_\eta}}{2\sqrt{\eta}}\left(\sum_{i=1}^{n}(\mathrm{Inf}_i^{\le d_\eta}(f))^2\right).$$

Proof. For any $\eta > 0$, we have

$$\begin{aligned}\vdash_6 \sum_{i=0}^{n-1}\mathop{\mathbf{E}}_{x\in\{-1,1\}^i}[\widehat{g}_x^2(n-i)\widehat{h}_x(n-i)\widehat{\ell}_x(n-i)] \le{}& \frac{1}{2\sqrt{\eta}}\left(\sum_{i=0}^{n-1}\mathop{\mathbf{E}}_{x\in\{-1,1\}^i}\widehat{h}_x^2(n-i)\right)\\ &+\frac{\sqrt{\eta}\left(\sum_{i=0}^{n-1}\mathbf{E}_{x\in\{-1,1\}^i}\widehat{g}_x^4(n-i)\widehat{\ell}_x^2(n-i)\right)}{2}.\end{aligned}$$

From the proof of Claim 6.12, we know that

$$A_p \vdash_3 \sum_{i=0}^{n-1}\mathop{\mathbf{E}}_{x\in\{-1,1\}^i}\widehat{h}_x^2(n-i) \le \eta.$$

Thus, we get

$$\vdash_6 \sum_{i=0}^{n-1}\mathop{\mathbf{E}}_{x\in\{-1,1\}^i}[\widehat{g}_x^2(n-i)\widehat{h}_x(n-i)\widehat{\ell}_x(n-i)] \le \frac{\sqrt{\eta}}{2} + \frac{\sqrt{\eta}\left(\sum_{i=0}^{n-1}\mathbf{E}_{x\in\{-1,1\}^i}\widehat{g}_x^4(n-i)\widehat{\ell}_x^2(n-i)\right)}{2}. \tag{6.15}$$

However,

$$\vdash_8 \sum_{i=0}^{n-1}\mathop{\mathbf{E}}_{x\in\{-1,1\}^i}\widehat{g}_x^4(n-i)\widehat{\ell}_x^2(n-i) \le \frac{\sum_{i=0}^{n-1}\mathbf{E}_{x\in\{-1,1\}^i}\widehat{g}_x^8(n-i) + \sum_{i=0}^{n-1}\mathbf{E}_{x\in\{-1,1\}^i}\widehat{\ell}_x^4(n-i)}{2}.$$

Following the same proof as in the proof of Claim 6.12, we can show that

$$A_p \vdash_9 \sum_{i=0}^{n-1} \mathop{\mathbf{E}}_{x\in\{-1,1\}^i} \widehat{g}_x^8(n-i) \le 1\,.$$

Similarly, from the argument in the proof of Claim 6.13, we can show that

$$A_p \vdash_4 \sum_{i=0}^{n-1} \mathop{\mathbf{E}}_{x\in\{-1,1\}^i} \widehat{\ell}_x^4(n-i) \le 9^{d_\eta} \cdot \left(\sum_{i=0}^{n-1} \left(\mathrm{Inf}_{n-i}^{\le d_\eta}(f)\right)^2\right).$$

Combining these, we get that

$$A_p \vdash_9 \sum_{i=0}^{n-1} \mathop{\mathbf{E}}_{x\in\{-1,1\}^i} \widehat{g}_x^4(n-i)\widehat{\ell}_x^2(n-i) \le \frac{1+9^{d_\eta}\cdot\left(\sum_{i=0}^{n-1}\left(\mathrm{Inf}_{n-i}^{\le d_\eta}(f)\right)^2\right)}{2}\,.$$

Plugging this in (6.15), we get the claim. □

Combining (6.12) with Claim 6.12, Claim 6.13, Claim 6.14 and plugging into Claim 6.3, we get that for c_γ and d_γ described in Lemma 6.6,

$$A_p \vdash_{\max\{d_\gamma,9\}} \mathop{\mathbf{E}}_{\substack{x\in\{-1,1\}^n\\ y\sim_{\rho'}x}}[\widetilde{J}(g(x),g(y))] \ge \widetilde{J}(\mathbf{E}[g(x)],\mathbf{E}[g(y)]) - \varepsilon - \frac{5\cdot c_\gamma\sqrt{\eta}}{2} - \frac{9^{d_\eta}\cdot c_\gamma}{\sqrt{\eta}}\Big(\sum_{i=1}^{n}(\mathrm{Inf}_i^{\le d_\eta}(f))^2\Big)\,. \tag{6.16}$$

Using Claim 6.5, we have that $A_p \vdash_{d_\alpha} g(x)\cdot g(y) \ge \widetilde{J}(g(x),g(y)) - 2\varepsilon$. Similarly, combining this with Claim 6.4 (and using that $d_\alpha \ge 2$), we get that

$$A_p \vdash_{d_\alpha} \mathbf{E}_{x,y\sim_\rho x}[f(x)\cdot f(y)] \ge \mathbf{E}_{x,y\sim_{\rho'}x}[\widetilde{J}(g(x),g(y))] - 4\varepsilon\,. \tag{6.17}$$

We now combine this with (6.16) to get that

$$A_p \vdash_{\max\{d_\gamma,d_\alpha,9\}} \mathop{\mathbf{E}}_{\substack{x\in\{-1,1\}^n\\ y\sim_{\rho}x}}[f(x)\cdot f(y)] \ge \widetilde{J}(\mathbf{E}[g(x)],\mathbf{E}[g(y)]) - 5\varepsilon - \frac{5\cdot c_\gamma\sqrt{\eta}}{2} - \frac{9^{d_\eta}\cdot c_\gamma}{\sqrt{\eta}}\Big(\sum_{i=1}^{n}(\mathrm{Inf}_i^{\le d_\eta}(f))^2\Big)\,. \tag{6.18}$$

Now, define a new sequence of indeterminates $\{f_c(x)\}_{x\in\{-1,1\}^n}$ where $f_c(x) = 1 - f(x)$. Analogous to the (sequence of) indeterminates $\{f_1(x)\}_{x\in\{-1,1\}^n}$ and $\{g(x)\}_{x\in\{-1,1\}^n}$ defined earlier, we define the indeterminates $\{f_{1c}(x)\}_{x\in\{-1,1\}^n}$ and $\{g_c(x)\}_{x\in\{-1,1\}^n}$ as follows:

$$f_{1c}(x) = (1-\varepsilon)f_c(x) + \varepsilon/2 \quad\text{ and }\quad g_c(x) = \mathbf{E}_{y\sim_{1-\eta}x}[f_{1c}(x)]\,.$$

We now make the following observations:

- $\forall x \in \{-1,1\}^n$, $A_p \vdash_1 \forall x \in \{-1,1\}^n$, $\varepsilon \le g_c(x) \le (1-\varepsilon)$.
- $\mathbf{E}_x[g(x)] + \mathbf{E}_x[g_c(x)] = 1$.
- For all $i \in [n]$, $\mathrm{Inf}_i^{\le d_\eta} f = \mathrm{Inf}_i^{\le d_\eta} f_c$.

Thus, using the above, analogous to (6.18), we have the following:

$$A_p \vdash_{\max\{d_\gamma, d_\alpha, 9\}} \mathop{\mathbf{E}}_{\substack{x \in \{-1,1\}^n \\ y \sim_\rho x}} [f_c(x) \cdot f_c(y)] \ge \widetilde{J}(\mathbf{E}[g_c(x)], \mathbf{E}[g_c(y)]) - 5\varepsilon \\ - \frac{5 \cdot c_\gamma \sqrt{\eta}}{2} - \frac{9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}} \Big(\sum_{i=1}^n (\mathrm{Inf}_i^{\le d_\eta}(f_c))^2 \Big). \tag{6.19}$$

Defining ξ as

$$\xi = 5\varepsilon + \frac{5 \cdot c_\gamma \sqrt{\eta}}{2} + \frac{9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}} \left(\sum_{i=1}^n (\mathrm{Inf}_i^{\le d_\eta}(f))^2 \right)$$

and summing up (6.18) and (6.19), we get

$$A_p \vdash_{\max\{d_\gamma, d_\alpha, 9\}} \mathop{\mathbf{E}}_{\substack{x \in \{-1,1\}^n \\ y \sim_\rho x}} [f(x) \cdot f(y) + (1-f(x)) \cdot (1-f(y))] \ge \widetilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) \\ + \widetilde{J}(\mathbf{E}[1-g(x)], \mathbf{E}[1-g(y)]) - 2\xi. \tag{6.20}$$

Next, we recall the following fact:

Fact 6.15. *For any $a \in (0,1)$ and $\rho \in (-1,0)$,*

$$J_\rho(a,a) + J_\rho(1-a, 1-a) \ge 2J_\rho(1/2, 1/2) = 1 - \frac{\arccos \rho}{\pi}.$$

As a consequence we have that for every $x \in [\varepsilon, 1-\varepsilon]$,

$$\widetilde{J}(x,x) + \widetilde{J}(1-x, 1-x) \ge 1 - \frac{\arccos \rho'}{\pi} - 2\varepsilon.$$

By using Corollary 5.13, we have that there exists $d_\delta = d_\delta(\varepsilon, \rho')$ such that

$$A_p \vdash_{d_\delta} \widetilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) + \widetilde{J}(\mathbf{E}[1-g(x)], \mathbf{E}[1-g(y)]) \ge 1 - \frac{\arccos \rho'}{\pi} - 4\varepsilon. \tag{6.21}$$

Combining (6.20) and (6.21), we get

$$A_p \vdash_{\max\{d_\gamma, d_\alpha, d_\delta, 9\}} \mathop{\mathbf{E}}_{\substack{x \in \{-1,1\}^n \\ y \sim_\rho x}} [f(x) \cdot f(y) + (1-f(x)) \cdot (1-f(y))] \ge 1 - \frac{\arccos \rho'}{\pi} - 4\varepsilon - 2\xi. \tag{6.22}$$

From here, getting to Theorem 6.1 is pretty easy. We first note that

$$4\varepsilon + 2\xi = 14\varepsilon + 5 \cdot c_\gamma \sqrt{\eta} + \frac{2 \cdot 9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}} \Big(\sum_{i=1}^n (\mathrm{Inf}_i^{\le d_\eta}(f_c))^2 \Big).$$

We now proceed as follows:

- For the given ρ and κ, first we choose $\varepsilon = \kappa/100$. This implies that $14 \cdot \varepsilon \leq \kappa/4$.
- Next, observe that $c_\gamma = c_\gamma(\rho', \varepsilon)$ is a continuous function of ρ' and ε. Now, recall that $\rho' = \rho/(1-\eta)$. At $\eta = 0$, $\sqrt{\eta} \cdot c_\gamma(\rho', \varepsilon) = 0$. Hence, there exists $\eta_0 = \eta_0(\rho, \varepsilon, \kappa) > 0$ such that for all $\eta \leq \eta_0$, $\sqrt{\eta} \cdot c_\gamma(\rho', \varepsilon) \leq \kappa/32$.
- Again, observe that for any $\rho \in (-1,0)$ and $\kappa > 0$, there exists $\eta_1 = \eta_1(\rho, \kappa) > 0$ such that for all $\eta \leq \eta_1$, $(\arccos \rho)/\pi \leq (\arccos \rho')/\pi + \kappa/4$.

Now, choose $\eta = \min\{\eta_0, \eta_1\}$. With η and ε having been fixed in terms of κ and ρ, we set

$$d_0(\kappa, \rho) = \max\{d_\gamma, d_\alpha, d_\delta, 9\}, \quad c(\kappa, \rho) = \frac{2 \cdot 9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}}, \quad \text{and} \quad d_1(\kappa, \rho) = d_\eta$$

and hence get

$$A_p \vdash_{d_0(\kappa,\rho)} \mathop{\mathbf{E}}_{\substack{x \in \{-1,1\}^n \\ y \sim_\rho x}} [f(x) \cdot f(y) + (1 - f(x)) \cdot (1 - f(y))] \geq 1 - \frac{\arccos \rho}{\pi} - \kappa$$
$$- c(\kappa, \rho) \cdot \left(\sum_{i=1}^{n} (\mathrm{Inf}_i^{\leq d_1(\kappa,\rho)}(f))^2 \right).$$

This finishes the proof of Theorem 6.1.

7 Refuting the Khot-Vishnoi instances of MAX-CUT

In this section, we will prove the following theorem:

Theorem 7.1. *Let $\rho \in (-1,0)$ and $G_\rho = (V_\rho, E_\rho)$ be the* MAX-CUT *instance constructed in* [33] *for the noise parameter ρ. Let $\{x_v\}_{v \in V}$ be a sequence of indeterminates and $A = \bigcup_{v \in V} \{0 \leq x_v \leq 1\}$. Then, for any $\delta > 0$, there exists $d' = d'(\delta, \rho)$ such that*

$$A \cup \left\{ \mathop{\mathbf{E}}_{(u,v) \in E} x_u \cdot (1 - x_v) + x_v \cdot (1 - x_u) \geq \frac{1}{\pi} \arccos \rho + \delta \right\} \vdash_{d'} -1 \geq 0.$$

Note that if the graph G_ρ has a cut with $\lambda \cdot |E|$ edges, then there is $\{0,1\}$ assignment to the set $\{x_v\}_{v \in V_\rho}$ such that $\mathbf{E}_{(u,v) \in E}\, x_u \cdot (1 - x_v) + x_v \cdot (1 - x_u) = \lambda$. Thus, Theorem 7.1 shows that for any $\delta > 0$, there is constant degree (dependent just on δ and ρ) SoS refutation for the assertion that the size of the fractional max-cut of G_ρ exceeds $(1/\pi) \arccos \rho + \delta$. This theorem is essentially tight as it is known that the size of max-cut in G_ρ is at least $(1/\pi) \arccos \rho - o(1)$. Thus, a constant number of rounds of the Lasserre hierarchy computes the value of max-cut in G_ρ (nearly) optimally.

To understand the significance of this result, we recall that O'Donnell and Zhou [49] had proved the same result with $(1/\pi) \arccos \rho + \delta$ replaced by

$$\frac{1}{2} - \frac{\rho}{\pi} - \left(\frac{1}{2} - \frac{1}{\pi} \right) \cdot \rho^3.$$

While this does show that a constant number of rounds of Lasserre does better than the basic SDP on the KV MAX-CUT instances, the question of whether with a constant number of rounds of Lasserre, the integrality gap on these instances can come arbitrarily close to 1 remained open. Theorem 7.1 answers this question in the affirmative.

The main technical component that O'Donnell and Zhou use in the proof of their analogue of Theorem 7.1 is the following theorem.

Theorem 7.2. [49] *Let* $\{f(x)\}_{x\in\{-1,1\}^n}$ *be a sequence of indeterminates and let* $A = \bigcup_{x\in\{-1,1\}^n}\{0 \le f(x) \le 1\}$. *Then, for any* $\delta > 0$ *and* $\rho \in (-1,0)$,

$$A \vdash_{O(1/\delta^2)} \mathrm{Stab}_\rho(f) \ge K(\rho) - \delta - 2^{O(1/\delta^2)} \cdot \left(\sum_{i=1}^{n} \widehat{f}^4(i)\right).$$

where

$$K(\rho) = \frac{1}{2} + \frac{\rho}{\pi} + \left(\frac{1}{2} - \frac{1}{\pi}\right) \cdot \rho^3 .$$

Further, they apply this theorem essentially in a black-box manner. The main idea behind strengthening the O'Donnell-Zhou result is that Theorem 6.1 is a strengthening of Theorem 7.2. Thus, we seek to follow the same steps as in [49] except we will use replace the application of Theorem 7.2 by Theorem 6.1. As Theorem 7.1 can be proved by just following the steps in [49] (and applying Theorem 6.1), we will not give its complete proof here. However, to help the reader, we do describe some of the important steps.

Unique Games. We begin by recalling the description of instances of UNIQUE-GAMES (UG) (see [30]). A UG instance consists of:

- an undirected graph $G = (V, E)$;
- a probability distribution $\mathcal{E}$ on tuples of the form $((u,v), \pi_{(u,v)})$ where $(u,v) \in E$ and $\pi_{(u,v)} : [k] \to [k]$ is a permutation.

Further, the weighted graph on G defined by $\mathcal{E}$ is regular. Also, $[k]$ is said to be the alphabet of the UNIQUE GAMES instance. The objective is to get a mapping $L : V \to [k]$ so as to maximize the following quantity:

$$\nu(G,\mathcal{E}) := \mathbf{Pr}_{(u,v,\pi_{(u,v)})\in\mathcal{E}}[\mathcal{L}(v) = \pi_{(u,v)}(\mathcal{L}(u))] .$$

The quantity $\nu(G,\mathcal{E})$ is said to be the value of the instance. Let $\mathcal{E}_u$ denote the marginal distribution on (v,π) when the first vertex is conditioned to be u. We next consider the SoS formulation for the UG instance described above. It is slightly different from the "obvious" formulation and follows the formulation in [49]. To motivate the formulation, we first state the following claim without proof. The proof can be found in [49].

Claim 7.3. *Let* $(G,\mathcal{E})$ *be a* UNIQUE GAMES *instance as defined above. Let* $\{x_{v,i}\}_{v\in V,\ i\in[k]}$ *be a set of real numbers satisfying the following conditions:*

- *For all* $v \in V$, $i \in [k]$, $x_{v,i} \ge 0$.

- *For all $v \in V$, $\sum_{i\in[k]} x_{v,i} \le 1$.*

If $\nu(G,\mathcal{E}) \le \beta$, then

$$\mathop{\mathbf{E}}_{u\in V}\left[\sum_{i=1}^{k}\left(\mathop{\mathbf{E}}_{(v,\pi_{(u,v)})\in\mathcal{E}_u} x_{v,\pi_{(u,v)}(i)}\right)^2\right] \le 4\beta\,.$$

Let $\{x_{v,i}\}_{v\in V,i\in[k]}$ be a set of indeterminates. Inspired by Claim 7.3, we make the following definition.

Definition 7.4. Given a UG instance $(G,\mathcal{E})$ with alphabet size k, there is a degree-d SOS refutation for optimum β if

$$A_p \bigcup \left\{\mathop{\mathbf{E}}_{u\in V}\left[\sum_{i=1}^{k}\left(\mathop{\mathbf{E}}_{(v,\pi_{u,v})\in\mathcal{E}_u} x_{v,\pi_{u,v}(i)}\right)^2\right] \ge \beta\right\} \vdash_d -1 \ge 0\,.$$

where

$$A_p = \bigcup_{v\in V,i\in[k]}\{x_{v,i}\ge 0\} \cup \bigcup_{v\in V}\left\{\sum_{i\in[k]} x_{v,i} \le 1\right\}.$$

Before we go ahead, we recall that for any $\eta\in(0,1)$ and $N\in\mathbb{N}$ (which is a power of 2), [33] construct UG instances over $2^N/N$ vertices, alphabet size n such that optimal value of the instance is bounded by $N^{-\eta}$.[2] Modifying the result from [3], O'Donnell and Zhou [49] show the following (Lemma 8.7 in [49]):

Theorem 7.5. *Let $\eta\in(0,1)$ and N be a power of 2 and let $(V,\mathcal{E})$ be the corresponding instances of UG constructed in [33]. Then, there is a degree-4 SoS refutation for optimum $\beta = N^{-\Omega(\eta)}$.*

We next describe the reduction from [30] of UG to MAX-CUT. The reduction is parameterized by a "correlation" value $\rho\in(-1,0)$. Given the instance of UG described above, the set of vertices in the corresponding MAX-CUT instance is given by $V' = V\times\{-1,1\}^k$. Further, the probability distribution $\mathcal{E}_{\rho,k}$ over the edges is given by the following sampling procedure:

- Choose $u \sim V$ uniformly at random.
- Choose $(u,v_1,\pi_{(u,v_1)})$ and $(u,v_2,\pi_{(u,v_2)})$ independently from the distribution $\mathcal{E}_u$.
- Choose $x\in\{-1,1\}^k$ and $y\sim_\rho x$.
- Output vertices $((v_1,\pi_{(u,v_1)}(x)),(v_2,\pi_{(u,v_2)}(y)))$.

We also have the following simple claim from [30]:

Claim 7.6. *Let $G' = (V',\mathcal{E}_{\rho,k})$ be an instance of* MAX-CUT *described above. Consider a partition of the graph G' (into two sets) specified by a collection of functions $\{f_v : \{-1,1\}^k \to \{0,1\}\}$. Then, the value of cut defined by this partition is $1-\mathbf{E}_{u\in V}[\mathrm{Stab}_\rho(g_u)]$ where*

$$g_u : \{-1,1\}^k \to [0,1] \quad \textit{is defined as} \quad g_u(x) = \mathop{\mathbf{E}}_{(v,\pi)\in\mathcal{E}_u}[f_v(\pi(x))]\,.$$

[2] Of course, the interesting part is that [33] shows that the standard SDP relaxation on this instance has value $1-\eta$

Having described the reduction from UNIQUE GAMES to MAX-CUT, we now consider the SoS relaxation of the MAX-CUT instance defined by V' and $\mathcal{E}_{\rho,k}$. In particular, we have an indeterminate $f_v(z)$ for every $v \in V$ and $z \in \{-1,1\}^k$. The constraint set A_m is given by

$$A_m = \bigcup_{v\in V} \bigcup_{z\in\{-1,1\}^k} \{0 \le f_v(z) \le 1\}\,.$$

Given Claim 7.6, we can say that there is a degree-d SoS refutation for the value of the max-cut in this graph exceeding β if

$$A_m \cup \{1 - \mathop{\mathbf{E}}_{u\in V}[\mathrm{Stab}_\rho(g_u)] \ge \beta\} \vdash_d -1 \ge 0.$$

Note that $\mathrm{Stab}_\rho(g_u)$ is defined in terms of the indeterminates $\{0 \le f_v(z) \le 1\}_{v\in V,\ z\in\{-1,1\}^k}$ in the obvious way (g_u is a homogeneous linear form in these indeterminates). The following claim (proven by O'Donnell and Zhou [49]) shows the connection between SoS refutation for UNIQUE GAMES instances and SoS refutation for MAX-CUT instances obtained after the [30] reduction.

Claim 7.7. *Let $(G,\mathcal{E})$ be a* UNIQUE GAMES *instance with a degree* 4 *SoS refutation for optimum β and $g(u)$ be the corresponding indeterminates obtained after the [30] reduction. Then,*

$$A_m \cup \left\{1 - \mathop{\mathbf{E}}_{u\in V}[\mathrm{Stab}_\rho(g_u)] \ge K(\rho) - \delta - 2^{O(1/\delta^2)}\beta\right\} \vdash_{O(1/\delta^2)+4} -1 \ge 0\,. \tag{7.1}$$

This implies that if the reduction from [30] is applied on the instances from Theorem 7.5,

$$A_m \cup \left\{1 - \mathop{\mathbf{E}}_{u\in V}[\mathrm{Stab}_\rho(g_u)] \ge K(\rho) - \delta - 2^{O(1/\delta^2)} \cdot N^{-\Omega(\eta)}\right\} \vdash_{O(1/\delta^2)+4} -1 \ge 0\,.$$

The proof of Claim 7.7 is reasonably straightforward and goes by proving SoS versions of many analytic statements and inequalities including use of Theorem 7.2. If we follow the exact same steps but use Theorem 6.1 instead of Theorem 7.2, we get the following claim.

Claim 7.8. *Let $(G,\mathcal{E})$ be a* UNIQUE GAMES *instance with a degree* 4 *SoS refutation for optimum β and $g(u)$ be the corresponding indeterminates obtained after the [30] reduction. Then,*

$$A_m \cup \left\{1 - \mathop{\mathbf{E}}_{u\in V}[\mathrm{Stab}_\rho(g_u)] \ge (\arccos\rho)/\pi - \delta - d_1(\delta,\rho)c(\delta,\rho)\cdot\beta\right\} \vdash_{d_0(\delta,\rho)+4} -1 \ge 0\,. \tag{7.2}$$

We do not repeat the steps here as it is identical to the proof of Claim 7.7. Using $\beta = N^{-\Omega(\eta)}$, we get Theorem 7.1.

Acknowledgements

We thank Ryan O'Donnell and Yuan Zhou for sharing the manuscript [49]. We thank the anonymous reviewers for their helpful comments.

A Facts regarding J_ρ

Here we collect various facts about the function

$$J_\rho(x,y) = \mathbf{Pr}[X \le \Phi^{-1}(x), Y \le \Phi^{-1}(y)]\,,$$

where

$$(X,Y) \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

As is standard, we will use ϕ to denote the density of the standard normal distribution. These calculations all follow from elementary calculus.

Claim 2.6. *For any $(x,y) \in (0,1)^2$ and $0 \le \sigma \le \rho$, $M_{\rho\sigma}(x,y)$ is a negative semidefinite matrix. Likewise, if $\rho \le \sigma \le 0$, then $M_{\rho\sigma}(x,y)$ is a positive semidefinite matrix.*

Proof. Towards proving this, note that we can define $Y = \rho \cdot X + \sqrt{1-\rho^2} \cdot Z$ where $Z \sim \mathcal{N}(0,1)$ is an independent normal. Also, let us define $\Phi^{-1}(x) = s$ and $\Phi^{-1}(y) = t$. For $s,t \in \mathbb{R}$, define $K_\rho(s,t)$ as

$$K_\rho(s,t) = \mathbf{Pr}_{X,Y}[X \le s, Y \le t] = \mathbf{Pr}_{X,Z}[X \le s, Z \le (t - \rho \cdot X)/\sqrt{1-\rho^2}]\,.$$

Note that for the aforementioned relations between x, y, s and t, $K_\rho(s,t) = J_\rho(x,y)$. Note that

$$K_\rho(s,t) = \int_{s'=-\infty}^{s} \phi(s') \int_{t'=-\infty}^{(t-\rho\cdot s')/\sqrt{1-\rho^2}} \phi(t') ds' dt'\,. \tag{A.1}$$

This implies that

$$\frac{\partial K_\rho(s,t)}{\partial s} = \phi(s) \int_{t'=-\infty}^{(t-\rho\cdot s)/\sqrt{1-\rho^2}} \phi(t') dt'\,.$$

By chain rule, we get that

$$\frac{\partial J_\rho(x,y)}{\partial x} = \frac{\partial K_\rho(s,t)}{\partial s} \cdot \frac{\partial s}{\partial x}\,.$$

By elementary calculus, it follows that

$$\frac{d\Phi^{-1}(x)}{dx} = \frac{1}{\phi(\Phi^{-1}(x))} \quad \Rightarrow \quad \frac{\partial s}{\partial x} = \frac{1}{\phi(\Phi^{-1}(x))} = \frac{1}{\phi(s)}\,.$$

Thus,

$$\frac{\partial J_\rho(x,y)}{\partial x} = \int_{t'=-\infty}^{(t-\rho\cdot s)/\sqrt{1-\rho^2}} \phi(t') dt'\,.$$

Thus, we next get that

$$\begin{aligned}
\frac{\partial^2 J_\rho(x,y)}{\partial x^2} &= \frac{\partial^2 J_\rho(x,y)}{\partial x \partial s} \cdot \frac{\partial s}{\partial x} = \phi\left(\frac{t - \rho \cdot s}{\sqrt{1-\rho^2}}\right) \cdot \frac{-\rho}{\sqrt{1-\rho^2}} \cdot \frac{1}{\phi(s)} \\
&= \phi\left(\frac{\Phi^{-1}(y) - \rho \cdot \Phi^{-1}(x)}{\sqrt{1-\rho^2}}\right) \cdot \frac{-\rho}{\sqrt{1-\rho^2}} \cdot \frac{1}{\phi(s)}\,. \\
\frac{\partial^2 J_\rho(x,y)}{\partial x \partial y} &= \frac{\partial^2 J_\rho(x,y)}{\partial x \partial t} \cdot \frac{\partial t}{\partial y} = \phi\left(\frac{\Phi^{-1}(y) - \rho \cdot \Phi^{-1}(x)}{\sqrt{1-\rho^2}}\right) \cdot \frac{1}{\sqrt{1-\rho^2}} \cdot \frac{1}{\phi(t)}\,.
\end{aligned}$$

Because we know that $(X,Y) \sim (Y,X)$, by symmetry, we can conclude that

$$\frac{\partial^2 J_\rho(x,y)}{\partial y^2} = \phi\left(\frac{\Phi^{-1}(x) - \rho\cdot\Phi^{-1}(y)}{\sqrt{1-\rho^2}}\right)\cdot\frac{-\rho}{\sqrt{1-\rho^2}}\cdot\frac{1}{\phi(t)},$$

and likewise,

$$\frac{\partial^2 J_\rho(x,y)}{\partial y\partial x} = \phi\left(\frac{\Phi^{-1}(x) - \rho\cdot\Phi^{-1}(y)}{\sqrt{1-\rho^2}}\right)\cdot\frac{1}{\sqrt{1-\rho^2}}\cdot\frac{1}{\phi(s)}.$$

It is obvious now that

$$\frac{\partial^2 J_\rho(x,y)}{\partial x^2}\cdot\frac{\partial^2 J_\rho(x,y)}{\partial y^2} - \rho^2\left(\frac{\partial^2 J_\rho(x,y)}{\partial x\partial y}\right)^2 = 0.$$

Now, suppose that $|\sigma| \le |\rho|$. Then

$$\det(M_{\rho\sigma}(x,y)) = \frac{\partial^2 J_\rho(x,y)}{\partial x^2}\cdot\frac{\partial^2 J_\rho(x,y)}{\partial y^2} - \sigma^2\left(\frac{\partial^2 J_\rho(x,y)}{\partial x\partial y}\right)^2 \ge 0.$$

If $\rho \ge 0$ then the diagonal of $M_{\rho\sigma}(x,y)$ is non-positive, and it follows that $M_{\rho\sigma}(x,y)$ is negative semidefinite. If $\rho \le 0$ then the diagonal is non-negative and so $M_{\rho\sigma}(x,y)$ is positive semidefinite. □

Claim 2.7. *For any $-1 < \rho < 1$, there exists $C(\rho) > 0$ such that for any $i, j \ge 0$, $i + j = 3$,*

$$\left|\frac{\partial^3 J_\rho(x,y)}{\partial x^i \partial y^j}\right| \le C(\rho)(xy(1-x)(1-y))^{-C(\rho)}.$$

Further, the function $C(\rho)$ can be chosen so that it is continuous for $\rho \in (-1,1)$.

Proof. As before, we set $\Phi^{-1}(x) = s$ and $\Phi^{-1}(y) = t$. From the proof of Claim 2.6, we see that

$$\frac{\partial^2 J_\rho(x,y)}{\partial x^2} = \phi\left(\frac{\Phi^{-1}(y) - \rho\cdot\Phi^{-1}(x)}{\sqrt{1-\rho^2}}\right)\cdot\frac{-\rho}{\sqrt{1-\rho^2}}\cdot\frac{1}{\phi(s)}.$$

To compute the third derivatives of J, recall that

$$\frac{\partial s}{\partial x} = \frac{1}{\phi(s)} \quad \text{and} \quad \frac{\partial t}{\partial y} = \frac{1}{\phi(t)},$$

we have

$$\begin{aligned}\frac{\partial^3 J_\rho(x,y)}{\partial x^3} &= \frac{\rho}{(1-\rho^2)^{3/2}}\frac{\rho t + (2\rho^2-1)s}{\phi(s)}\exp\left(-\frac{t^2 - 2\rho st + (2\rho^2-1)s^2}{2(1-\rho^2)}\right)\\ &= \frac{\sqrt{2\pi}\rho}{(1-\rho^2)^{3/2}}(\rho t + (2\rho^2-1)s)\exp\left(-\frac{t^2 - 2\rho st + (3\rho^2-2)s^2}{2(1-\rho^2)}\right).\end{aligned} \tag{A.2}$$

Now, $\Phi^{-1}(x) \sim \sqrt{2\log x}$ as $x \to 0$; hence there is a constant C such that $\Phi^{-1}(x) \le C\sqrt{\log x}$ for all $x \le 1/2$. Hence, $\exp(s^2) \le x^{-C}$ for all $x \le 1/2$; by symmetry, $\exp(s^2) \le (x(1-x))^{-C}$ for all $x \in (0,1)$. Therefore

$$\begin{aligned}
\exp\Big(-\frac{t^2-2\rho st+(3\rho^2-2)s^2}{2(1-\rho^2)}\Big) &= e^{-\frac{t^2}{2(1-\rho^2)}} e^{\frac{\rho st}{1-\rho^2}} e^{\frac{(2-3\rho^2)s^2}{2(1-\rho^2}} \\
&\le e^{-\frac{t^2}{2(1-\rho^2)}} e^{\frac{\rho(s^2+t^2)}{2(1-\rho^2)}} e^{\frac{(2-3\rho^2)s^2}{2(1-\rho^2}} \\
&\le \big(x(1-x)y(1-y)\big)^{-\frac{\rho}{2(1-\rho^2)}} \big(x(1-x)\big)^{-\frac{2-3\rho^2}{2(1-\rho^2)}} .
\end{aligned} \tag{A.3}$$

Further, using $\exp(s^2) \le x^{-C}$ and $\exp(t^2) \le y^{-C}$, $\rho t + (2\rho^2-1)s \le 4(xy)^{-C}$. As a consequence, applying this to (A.2), we see that there is a constant $C(\rho) > 0$,

$$\left|\frac{\partial^3 J_\rho(x,y)}{\partial x^3}\right| \le C(\rho)\big(x(1-x)y(1-y)\big)^{-C(\rho)}.$$

The other third derivatives are similar:

$$\frac{\partial^3 J_\rho(x,y)}{\partial x^2 \partial y} = \frac{\sqrt{2\pi}\rho}{(1-\rho^2)^{3/2}}(t-2\rho s)\exp\Big(-\frac{(2\rho^2-1)t^2-2\rho st+(2\rho^2-1)s^2}{2(1-\rho^2)}\Big).$$

By the same steps that led to (A.3), we get

$$\left|\frac{\partial^3 J_\rho(x,y)}{\partial x^2 \partial y}\right| \le C(\rho)\big(x(1-x)y(1-y)\big)^{-C(\rho)}$$

(for a slightly different $C(\rho)$). The bounds on $\partial^3 J/\partial y^2 \partial x$ and $\partial^3 J/\partial x^3$ then follow because J is symmetric in x and y.

The fact that $C(\rho)$ can be chosen so that it is continuous for $\rho \in (-1,1)$ is obvious from the discussion above. □

Claim A.1. *For any $x, y \in (0,1)$,*

$$\left|\frac{\partial J_\rho(x,y)}{\partial \rho}\right| \le (1-\rho^2)^{-3/2}.$$

Proof. We begin from (A.1), but this time we differentiate with respect to ρ:

$$\frac{\partial K_\rho(s,t)}{\partial \rho} = -\frac{1}{(1-\rho^2)^{3/2}} \int_{s'=-\infty}^{s} \phi(s')\phi\left(\frac{t-\rho s'}{\sqrt{1-\rho^2}}\right) ds' .$$

Since $Range(\phi) \subset (0,1]$ and $\int_{s'} \phi(s')ds' = 1$, it follows that

$$\left|\frac{\partial K_\rho(s,t)}{\partial \rho}\right| \le (1-\rho^2)^{-3/2}.$$

Since

$$\frac{\partial J_\rho(s,t)}{\partial \rho} = \frac{\partial K_\rho(\Phi^{-1}(x),\Phi^{-1}(y))}{\partial \rho},$$

the proof is complete. □

We also state the following useful claim without a proof. The proof is obvious from the calculations in the proofs of Claim 2.6 and Claim 2.7.

Claim A.2. *For any $\rho \in (-1,1)$, $\varepsilon > 0$ there exists a continuous function $\gamma(\rho,\varepsilon)$ such that for any $(x,y) \in [\varepsilon, 1-\varepsilon]^2$ and $1 \le i+j \le 3$,*

$$\left| \frac{\partial^{i+j} J_\rho(x,y)}{\partial^i x \partial^j y} \right| \le \gamma(\rho,\varepsilon)\,.$$

B Approximation by polynomials

Claim B.1. *For any $\rho \in (-1,1)$ and any $\delta > 0$, there is a polynomial $\tilde{J}$ such that for all $0 \le i+j \le 2$,*

$$\sup_{x,y\in[\varepsilon,1-\varepsilon]} \left| \frac{\partial^{i+j} J_\rho(x,y)}{\partial x^i \partial y^j} - \frac{\partial^{i+j} \tilde{J}_\rho(x,y)}{\partial x^i \partial y^j} \right| \le \delta\,.$$

Moreover, if $\rho \in [-1+\varepsilon, 1-\varepsilon]$, then the degree of $\tilde{J}$ and the maximal coefficient in $\tilde{J}$ can be bounded by constants depending only on ε and δ. Further, both these constants are continuous functions of ε and δ for any $\varepsilon > 0$.

The proof of Claim B.1 follows from standard results on Bernstein polynomials. In particular, we make use of the following theorem which may be found, for example, in [39].

Theorem B.2. *Suppose $f : [0,1] \to \mathbb{R}$ has m continuous derivatives which are all bounded in absolute value by M. For any $n \in \mathbb{N}$, let $B_n f$ be the polynomial*

$$(B_n f)(x) = \sum_{k=1}^{n} f(k/n) \binom{n}{k} x^k (1-x)^{n-k}\,.$$

Then for any $0 \le i \le m$,

$$\sup_{x\in[0,1]} \left| \frac{d^i f(x)}{dx^i} - \frac{d^i (B_n f)(x)}{dx^i} \right| \le C\sqrt{M/n}\,.$$

Seeing as the first three derivatives of J_ρ are bounded on $[\varepsilon, 1-\varepsilon]$, Claim B.1 is essentially just a 2-variable version of Theorem B.2. Although such a result is almost certainly known (and for more than 2 variables), we were not to find a reference in the literature, and so we include the proof here.

Proof of Claim B.1. Suppose that $f : [0,1]^2 \to \mathbb{R}$ has all partial derivatives up to third order bounded by M. Define

$$g_n(x,y) = (B_n f(\cdot,y))(x) = \sum_{k=1}^{n} \binom{n}{k} f(k/n,y) x^k (1-x)^{n-k}\,.$$

$$h_n(x,y) = (B_n g_n(x,\cdot))(y) = \sum_{k=1}^{n} \sum_{\ell=1}^{n} \binom{n}{k} \binom{n}{\ell} f(k/n,\ell/n) x^k (1-x)^{n-k} y^\ell (1-y)^{n-\ell}\,.$$

Fix $0 \le i + j \le 2$ and note that

$$\frac{\partial^j g_n(\cdot, y)}{\partial y^j} = B_n \frac{\partial^j f(\cdot, y)}{\partial y^j}, \tag{B.1}$$

$$\frac{\partial^i h_n(x, \cdot)}{\partial x^i} = B_n \frac{\partial^i g_n(x, \cdot)}{\partial x^i}. \tag{B.2}$$

Now, fix $y \in [0,1]$ and apply Theorem B.2 to (B.1): for any $x \in [0,1]$,

$$\left| \frac{\partial^{i+j} g_n(x,y)}{\partial x^i \partial y^j} - \frac{\partial^{i+j} f(x,y)}{\partial x^i \partial y^j} \right| \le C\sqrt{M/n}.$$

On the other hand, fixing x and applying Theorem B.2 to (B.2) yields

$$\left| \frac{\partial^{i+j} h_n(x,y)}{\partial x^i \partial y^j} - \frac{\partial^{i+j} g_n(x,y)}{\partial x^i \partial y^j} \right| \le C\sqrt{M/n}.$$

Putting these together,

$$\left| \frac{\partial^{i+j} h_n(x,y)}{\partial x^i \partial y^j} - \frac{\partial^{i+j} f(x,y)}{\partial x^i \partial y^j} \right| \le 2C\sqrt{M/n}.$$

Since h_n is a polynomial, taking n sufficiently large implies that there is a polynomial $\tilde{f}$ such that $\tilde{f}$, and its partial derivatives of order at most 2, uniformly approximate the corresponding derivatives of f. Although we stated this for functions on $[0,1]^2$, a change of coordinates shows that it holds equally well for functions on $[\delta, 1-\delta]^2$ with three bounded derivatives. Since J_ρ is such a function, the first part of the claim follows.

For the second part of the claim, note that all of the error bounds hold uniformly in $\rho \in [-1+\varepsilon, 1-\varepsilon]$ since the first, second and third derivatives of J_ρ are uniformly bounded over $\rho \in [-1+\varepsilon, 1-\varepsilon]$ (follows easily by Claim A.2). Moreover, since $\max_{x,y} |J_\rho(x,y)| \le 1$, the coefficients in h_n can be bounded in terms of n, which is in turn bounded in terms of ε and δ.
Further, note that since the first, second and third derivatives derivatives of J_ρ are bounded by a continuous function of ρ and ε (see Claim A.2), hence n can also be bounded by a continuous function of ε and δ. As $\max_{x,y} |J_\rho(x,y)| \le 1$, the coefficients of $\tilde{J}_\rho$ are bounded by a continuous function of ε and δ and so is the degree of $\tilde{J}_\rho$. □

References

[1] PER AUSTRIN: Balanced MAX-2-SAT might not be the hardest. In *Proc. 39th STOC*, pp. 189–197. ACM Press, 2007. Available at ECCC. [doi:10.1145/1250790.1250818] 4

[2] DOMINIQUE BAKRY AND MICHEL LEDOUX: Lévy-Gromov's isoperimetric inequality for an infinite-dimensional diffusion generator. *Inventiones Mathematicae*, 123(1):259–281, 1996. [doi:10.1007/BF01232376] 4

[3] BOAZ BARAK, FERNANDO G. S. L. BRANDÃO, ARAM WETTROTH HARROW, JONATHAN A. KELNER, DAVID STEURER, AND YUAN ZHOU: Hypercontractivity, sum-of-squares proofs, and their applications. In *Proc. 44th STOC*, pp. 307–326. ACM Press, 2012. [doi:10.1145/2213977.2214006] 2, 6, 7, 19, 21, 39

[4] BOAZ BARAK, PRASAD RAGHAVENDRA, AND DAVID STEURER: Rounding semidefinite programming hierarchies via global correlation. In *Proc. 52nd FOCS*, pp. 472–481. IEEE Comp. Soc. Press, 2011. [doi:10.1109/FOCS.2011.95] 6

[5] WILLIAM BECKNER: Inequalities in Fourier analysis. *Ann. of Math.*, 102(1):159–182, 1975. [doi:10.2307/1970980] 3, 6, 15

[6] SERGEY G. BOBKOV: An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. *Ann. Probab.*, 25(1):206–214, 1997. Available at Project Euclid. 4

[7] ALINE BONAMI: Étude des coefficients Fourier des fonctions de $L^p(G)$. *Annales de l'Institute Fourier*, 20(2):335–402, 1970. EuDML. 3, 6, 15

[8] CHRISTER BORELL: Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Probab. Theory and Related fields*, 70(1):1–13, 1985. [doi:10.1007/BF00532234] 1, 3, 7, 8

[9] JEAN BOURGAIN: On the distributions of the Fourier spectrum of Boolean functions. *Israel J. Math.*, 131(1):269–276, 2002. [doi:10.1007/BF02785861] 2, 3

[10] EDEN CHLAMTAC AND GYANIT SINGH: Improved approximation guarantees through higher levels of SDP hierarchies. In *Proc. 11th Internat. Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'08)*, volume 5171 of *LNCS*, pp. 49–62. Springer, 2008. [doi:10.1007/978-3-540-85363-3_5] 6

[11] ANINDYA DE, ELCHANAN MOSSEL, AND JOE NEEMAN: Majority is stablest: discrete and SoS. In *Proc. 45th STOC*, pp. 477–486. ACM Press, 2013. [doi:10.1145/2488608.2488668, arXiv:1211.1001] 1

[12] NIKHIL R. DEVANUR, SUBHASH KHOT, RISHI SAKET, AND NISHEETH K. VISHNOI: Integrality gaps for sparsest cut and minimum linear arrangement problems. In *Proc. 38th STOC*, pp. 537–546. ACM Press, 2006. [doi:10.1145/1132516.1132594] 6

[13] IRIT DINUR, ELCHANAN MOSSEL, AND ODED REGEV: Conditional hardness for approximate coloring. *SIAM J. Comput.*, 39(3):843–873, 2009. Preliminary version in STOC'06. [doi:10.1137/07068062X] 4

[14] IRIT DINUR AND SAMUEL SAFRA: On the hardness of approximating minimum vertex-cover. *Ann. of Math.*, 162(1):439–485, 2005. Preliminary version in STOC'02. [doi:10.4007/annals.2005.162.439] 2

[15] WILLIAM FELLER: *An Introduction to Probability Theory and its Applications*. John Wiley & Sons, 1968. 12

[16] DAN S. FELSENTHAL AND MOSHÉ MACHOVER: *The Measurement of Voting Power*. Edward Elgar Publishing, 1998. 2

[17] EHUD FRIEDGUT AND GIL KALAI: Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124(10):2993–3002, 1996. [doi:10.1090/S0002-9939-96-03732-X] 2, 3

[18] EHUD FRIEDGUT, GIL KALAI, AND NOAM NISAN: Elections can be manipulated often. In *Proc. 49th FOCS*, pp. 243–249. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.87] 4

[19] MICHEL X. GOEMANS AND DAVID P. WILLIAMSON: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995. Preliminary version in STOC'94. [doi:10.1145/227683.227684] 3

[20] DIMA GRIGORIEV: Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoret. Comput. Sci.*, 259(1-2):613–622, 2001. [doi:10.1016/S0304-3975(00)00157-2] 5

[21] DIMA GRIGORIEV AND NICOLAI VOROBJOV: Complexity of Null- and Positivstellensatz proofs. *Ann. Pure Appl. Logic*, 113(1-3):153–160, 2001. [doi:10.1016/S0168-0072(01)00055-0] 5

[22] JOHAN HÅSTAD: Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001. Preliminary version in STOC'97. [doi:10.1145/502090.502098] 2

[23] DAVID HILBERT: Über die Darstellung definiter Formen als Summe von Formenquadraten. *Mathematische Annalen*, 32(3):342–350, 1888. [doi:10.1007/BF01443605] 5

[24] MARCUS ISAKSSON AND ELCHANAN MOSSEL: Maximally stable Gaussian partitions with discrete applications. *Israel J. Math.*, 189(1):347–396, 2012. [doi:10.1007/s11856-011-0181-7, arXiv:0903.3362] 4

[25] VAITHILINGAM JEYAKUMAR, JEAN BERNARD LASSERRE, AND GUOYIN LI: On polynomial optimization over non-compact semi-algebraic sets. *J. Optim. Theory and Appl.*, 163(3):707–718, 2014. [doi:10.1007/s10957-014-0545-3] 5, 21, 22

[26] JEFF KAHN, GIL KALAI, AND NATHAN LINIAL: The influence of variables on boolean functions (extended abstract). In *Proc. 29th FOCS*, pp. 68–80. IEEE Comp. Soc. Press, 1988. [doi:10.1109/SFCS.1988.21923] 2, 3, 6

[27] GIL KALAI: A Fourier-theoretic perspective on the Condorcet paradox and Arrow's theorem. *Adv. in Appl. Math.*, 29(3):412–426, 2002. [doi:10.1016/S0196-8858(02)00023-4] 1, 3

[28] GIL KALAI: Social indeterminacy. *Econometrica*, 72(5):1565–1581, 2004. [doi:10.1111/j.1468-0262.2004.00543.x] 2

[29] SUBHASH KHOT: On the power of Unique 2-Prover 1-Round Games. In *Proc. 34th STOC*, pp. 767–775. ACM Press, 2002. [doi:10.1145/509907.510017] 2

[30] SUBHASH KHOT, GUY KINDLER, ELCHANAN MOSSEL, AND RYAN O'DONNELL: Optimal inapproximability results for Max-Cut and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007. Preliminary version in FOCS'04. [doi:10.1137/S0097539705447372] 1, 2, 3, 6, 7, 38, 39, 40

[31] SUBHASH KHOT, PREYAS POPAT, AND RISHI SAKET: Approximate Lasserre integrality gap for Unique Games. In *Proc. 13th Internat. Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'10)*, volume 6302 of *LNCS*, pp. 298–311. Springer, 2010. [doi:10.1007/978-3-642-15369-3_23] 6

[32] SUBHASH KHOT AND RISHI SAKET: SDP integrality gaps with local ℓ_1-embeddability. In *Proc. 50th FOCS*, pp. 565–574. IEEE Comp. Soc. Press, 2009. [doi:10.1109/FOCS.2009.37] 6

[33] SUBHASH KHOT AND NISHEETH K. VISHNOI: The Unique Games Conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. *J. ACM*, 62(1), 2015. Preliminary version in FOCS'05. [doi:10.1145/2629614, arXiv:1305.4581] 2, 6, 7, 37, 39

[34] GUY KINDLER AND RYAN O'DONNELL: Gaussian noise sensitivity and Fourier tails. In *Proc. 27th IEEE Conf. on Computational Complexity (CCC'12)*, pp. 137–147. IEEE Comp. Soc. Press, 2012. [doi:10.1109/CCC.2012.35] 4

[35] JEAN-LOUIS KRIVINE: Anneaux préordonnés. *J. Analyse Math.*, 12:307–326, 1964. Available at HAL. 5

[36] JEAN BERNARD LASSERRE: Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2001. [doi:10.1137/S1052623400366802] 2, 5, 6, 19

[37] JEAN BERNARD LASSERRE: *Moments, Positive Polynomials and Their Applications*. Imperial College Press, London, 2010. [doi:10.1142/9781848164468_fmatter] 7, 22

[38] HENRI LOMBARDI, NIKOLAI MNEV, AND MARIE-FRANÇOISE ROY: The Positivstellensatz and small deduction rules for systems of inequalities. *Mathematische Nachrichten*, 181(1):245–259, 1996. [doi:10.1002/mana.3211810110] 5

[39] GEORGE G. LORENTZ: *Bernstein polynomials*. Chelsea Publishing Company, Inc., 1986. 6, 44

[40] LÁSZLÓ LOVÁSZ AND ALEXANDER SCHRIJVER: Cones of matrices and set-functions and 0-1 optimization. *SIAM J. Optim.*, 1(2):166–190, 1991. [doi:10.1137/0801013] 5, 6

[41] ELCHANAN MOSSEL: Lecture notes on Fourier analysis. Available at author's website, 2005. 13

[42] ELCHANAN MOSSEL: Gaussian bounds for noise correlation of functions. *Geom. Funct. Anal.*, 19(6):1713–1756, 2010. Preliminary version in FOCS'08. [doi:10.1007/s00039-010-0047-x, arXiv:math/0703683] 4, 17

[43] ELCHANAN MOSSEL: A quantitative Arrow theorem. *Probab. Theory and Related Fields*, 154(1–2):49–88, 2012. [doi:10.1007/s00440-011-0362-7, arXiv:0903.2574] 4

[44] ELCHANAN MOSSEL AND JOE NEEMAN: Robust optimality of Gaussian noise stability. *J. Eur. Math. Soc.*, 17(2):433–482, 2015. [doi:10.4171/JEMS/507, arXiv:1210.4126] 4, 7, 8, 9, 11

[45] ELCHANAN MOSSEL, RYAN O'DONNELL, AND KRZYSZTOF OLESZKIEWICZ: Noise stability of functions with low influences: invariance and optimality. *Ann. of Math.*, 171(1):295–341, 2010. Preliminary version in FOCS'05. [doi:10.4007/annals.2010.171.295] 1, 3, 6, 7, 23

[46] YURII NESTEROV: *Global quadratic optimization via conic relaxation*. Handbook of Semidefinite Programming. Kluwer, 2000. Available from Citeseer. 5

[47] RYAN O'DONNELL: *Analysis of Boolean Functions*. Cambridge University Press, 2014. 13

[48] RYAN O'DONNELL AND YI WU: An optimal SDP algorithm for Max-Cut, and equally optimal long code tests. In *Proc. 40th STOC*, pp. 335–344. ACM Press, 2008. [doi:10.1145/1374376.1374425] 4

[49] RYAN O'DONNELL AND YUAN ZHOU: Approximability and proof complexity. In *Proc. 24th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'13)*, pp. 1537–1556. ACM Press, 2013. [doi:10.1137/1.9781611973105.111] 4, 6, 7, 19, 20, 21, 26, 37, 38, 39, 40

[50] PABLO A. PARRILO: *Structured Semidefinite Programs and Semialgebraic Methods in Robustness and Optimization*. Ph. D. thesis, California Institute of Technology, 2000. Available from the Caltech Thesis Library. 5, 6, 19

[51] PRASAD RAGHAVENDRA AND DAVID STEURER: Integrality gaps for strong SDP relaxations of UNIQUE GAMES. In *Proc. 50th FOCS*, pp. 575–585. IEEE Comp. Soc. Press, 2009. [doi:10.1109/FOCS.2009.73] 6

[52] ALEX SAMORODNITSKY AND LUCA TREVISAN: Gowers uniformity, influence of variables, and PCPs. *SIAM J. Comput.*, 39(1):323–360, 2009. Preliminary version in STOC'06. [doi:10.1137/070681612, arXiv:math/0510264] 2

[53] GRANT SCHOENEBECK: Linear level Lasserre lower bounds for certain k-CSPs. In *Proc. 49th FOCS*, pp. 593–602. IEEE Comp. Soc. Press, 2008. [doi:10.1109/FOCS.2008.74] 5

[54] WILLIAM FLEETWOOD SHEPPARD: On the application of the theory of error to cases of normal distribution and normal correlation. *Phil. Trans. Royal Soc. London*, 192:101–167, 1899. [doi:10.1098/rsta.1899.0003] 3

[55] NAUM Z. SHOR: Class of global minimum bounds of polynomial functions. *Cybernetics*, 23(6):731–734, 1987. [doi:10.1007/BF01070233] 5

[56] GILBERT STENGLE: A Nullstellensatz and a Positivstellensatz in semialgebraic geometry. *Mathematische Annalen*, 207(2):87–97, 1974. Available at EuDML. [doi:10.1007/BF01362149] 5

[57] MICHEL TALAGRAND: On Russo's approximate 0-1 law. *Ann. Probab.*, 22(3):1576–1587, 1994. Available from Project Euclid. 2, 3

AUTHORS

Anindya De
Assistant Professor
Northwestern University
anindya@eecs.northwestern.edu
http://users.eecs.northwestern.edu/~anindya/

Elchanan Mossel
Professor
Massachusetts Institute of Technology
mossel@stat.berkeley.edu
http://www.stat.berkeley.edu/~mossel/

Joe Neeman
Assistant Professor
University of Texas, Austin
neeman@iam.uni-bonn.de
https://wt.iam.uni-bonn.de/neeman/

ABOUT THE AUTHORS

ANINDYA DE is an Assistant Professor of Computer Science at Northwestern University. He received his Ph. D. from U. C. Berkeley in 2013 advised by Luca Trevisan and Umesh Vazirani. From 2013 to 2015, he was a postdoc at the Institute for Advanced Study and DIMACS. His research interests are in complexity theory, learning theory, discrete analysis and applied probability.

ELCHANAN MOSSEL is a Professor of Mathematics at MIT. He received his Ph. D. in mathematics from Hebrew University in 2000, advised by Yuval Peres. After his Ph. D., he was a post-doc at the Theory Group of Microsoft Research, Redmond and a Miller fellow in Statistics and Computer Science at U. C. Berkeley. His research interests include Combinatorial Statistics, Discrete Fourier Analysis and Influences, Randomized Algorithms, Computational Complexity, MCMC, Markov Random Fields, Social Choice, Game Theory and Evolution.

JOE NEEMAN is a Professor in Mathematics at the University of Bonn and an Assistant Professor of Mathematics at the University of Texas, Austin. He received his Ph. D. from U.C. Berkeley, advised by Elchanan Mossel, after which he was a postdoc at U. T. Austin. He works in various areas of probability and on applications to computer science.