



University of Pennsylvania
ScholarlyCommons

Statistics Papers

Wharton Faculty Research

2013

Majority Is Stablest: Discrete and SoS

Anindya De

University of California - Berkeley

Elchanan Mossel

University of Pennsylvania

Joe Neeman

University of California - Berkeley

Follow this and additional works at: http://repository.upenn.edu/statistics_papers

 Part of the [Statistics and Probability Commons](#), and the [Theory and Algorithms Commons](#)

Recommended Citation

De, A., Mossel, E., & Neeman, J. (2013). Majority Is Stablest: Discrete and SoS. *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing, STOC '13* 477-486. <http://dx.doi.org/10.1145/2488608.2488668>

At the time of publication, author Elchanan Mossel was affiliated with the University of California, Berkeley. Currently, he is a faculty member at the Statistic Department at the University of Pennsylvania.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/statistics_papers/196

For more information, please contact repository@pobox.upenn.edu.

Majority Is Stablest: Discrete and SoS

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. The new proof is general enough to include all previous variants of majority is stablest such as "it ain't over until it's over" and "Majority is most predictable". 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 Khot-Vishnoi instance of Max-Cut does not provide a gap instance for the Lasserre hierarchy.

Keywords

majority is stablest, sum of squares hierarchy, unique games conjecture

Disciplines

Statistics and Probability | Theory and Algorithms

Comments

At the time of publication, author Elchanan Mossel was affiliated with the University of California, Berkeley. Currently, he is a faculty member at the Statistic Department at the University of Pennsylvania.

Majority is Stablest : Discrete and SoS

Anindya De*

University of California, Berkeley

Elchanan Mossel†

University of California, Berkeley

Joe Neeman‡

University of California, Berkeley

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. The new proof is general enough to include all previous variants of majority is stablest such as "it ain't over until it's over" and "Majority is most predictable". 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 Khot-Vishnoi instance of Max-Cut does not provide a gap instance for the Lasserre hierarchy.

arXiv:1211.1001v2 [cs.CC] 6 Nov 2012

*anindya@cs.berkeley.edu. Research supported by Umesh Vazirani's Templeton Foundation Grant 21674.

†mossel@stat.berkeley.edu. Research supported by NSF award DMS-1106999 and DOD ONR grant N000141110140

‡jneeman@stat.berkeley.edu. Research supported by NSF award DMS-1106999 and DOD ONR grant N000141110140

1 Introduction

The proof of the Majority is Stablest Theorem [MOO10] affirmed a conjecture in hardness of approximation [KKMO07] and in social choice [Kal02]. The result has been since extensively used in the two areas. One of the surprising features of the proof of [MOO10] is the crucial use of deep results in Gaussian analysis [Bor85] and an ‘‘Invariance Principle’’ that allows to deduce the discrete result from the Gaussian one.

Since the statement of the Majority is Stablest Theorem [MOO10] 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 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’’.

We also show how the new proof can be transformed into a ‘‘Sum of Squares’’ proof of the Majority is Stablest Theorem, thus showing that Khot-Vishnoi instance of Max-cut [KV05] does not provide an integrality gap instance for Max-cut in the Lasserre hierarchy.

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

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

where \mathbf{P} denotes 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 [KKL88, Tal94, FK96], the use of hyper-contractive estimates applied to low influence variables is one of the main techniques in discrete Fourier analysis.

Of particular interest are functions all of whose influences are low. The work of Friedgut and Kalai [FK96] shows that low influence functions have sharp thresholds. Central work in theoretical computer science [Bou02, DS05, ST06] pointed to the importance of low influence functions, including in the context of the ‘‘Unique Games Conjecture’’ [Kho02, KKMO07]. Such functions have also attracted much interest in the theory of social choice, see e.g. [FM98, Kal04].

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 who have a strong dependency on a small number of variables, see e.g. [Hås97, Kho02, DS05].

The Majority is Stablest theorem has been crucial in developments in both hardness of approximation and the theory of social choice. The 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 $\mathbf{E}[f(x)f(y) + (1 - f(x))(1 - f(y))]$ is a dictator function. The majority is stablest theorem states that for functions with low influences the value of $\mathbf{E}[f(x)f(y) + (1 - f(x))(1 - f(y))]$ 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 $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ at ρ is defined to be

$$\text{Stab}_\rho(f) := \mathbf{E}[f(x)f(y) + (1 - f(x))(1 - f(y))],$$

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 $\mathbf{E}[x_i y_i] = \rho$.

Theorem 1.2. ‘‘Majority Is Stablest’’ [MOO10] Let $0 \leq \rho \leq 1$ and $\epsilon > 0$ be given. Then there exists $\tau > 0$ such that if $f : \{-1, 1\}^n \rightarrow [0, 1]$ satisfies $\mathbf{E}[f] = 1/2$ and $\text{Inf}_i(f) \leq \tau$ for all i , then

$$\text{Stab}_\rho(f) \leq 1 - \frac{\arccos \rho}{\pi} + \epsilon.$$

By Sheppard’s Formula [She99], the quantity $1 - \frac{\arccos \rho}{\pi}$ is precisely $\lim_{n \rightarrow \infty} \text{Stab}_\rho(\text{Maj}_n)$, where

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

We also remark here that Theorem 1.2 readily generalizes to the case when $\mathbf{E}[f] \neq 1/2$ with the right hand side replaced by the corresponding quantity for a shifted majority with the same measure. This statement of Majority

is Stablest was conjectured in [KKMO07] 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’ 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 [GW95] 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 [Kal02]. The results of [MOO10] 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 [Kal02, MOO10] for more details.

The *statement* of the theorem deals with Boolean functions, yet *the proof* of [MOO10] crucially relies on Gaussian analysis as (a) it uses a deep result (with a hard proof) of Borell [Bor85] on noise stability in Gaussian space and (b) it uses the invariance principle developed in [MOO10] 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 [Bon70, Bec75] are proved by induction on dimension in the discrete cube. Moreover, most of the results in the area starting from KKL including [KKL88, Tal94, FK96, Bou02] 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

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 [MN12] 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 [Bob97]. Bobkov proved by induction a discrete functional inequality that at the limit becomes the Gaussian isoperimetric inequality. Moreover, Bobkov’s functional is crucial for the semi-group proof of the Gaussian isoperimetric inequality. In [MN12] a functional version of Borell’s result is defined and proved using the “semi-group” method in Gaussian space. Here we prove a discrete version of the same functional inequality.

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 proof of the result including using spherical symmetrization, see e.g. [IM12], sub-additivity [KO12] and a semi-group proof [MN12]. It is the simplest proof of Borell’s result using elementary arguments only ([IM12] uses sophisticated spherical re-arrangement inequalities, [KO12] only works for sets of measure 1/2 and certain noise values and [MN12] requires basic facts on the Orenstein-Uhlenbeck process).

Since it was proved, Theorem 1.2 was generalized a number of times including in [DMR06, Mos10]. The results and their generalization have been used numerous times in hardness of approximation and social choice including in [Aus07, OW08, Rag08, Mos12, FKN08]. Our simple proof extends to cover all of the generalization above. It also enables to prove an SoS version of the statement of Majority is Stablest, thus answering the main open problem of [OZ13] 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 (introduced by Grigoriev and Vorobjov [GV01]) 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 $X = (X_1, \dots, X_n)$ be a sequence of variables and let $\mathbb{R}[X]$ be the ring of polynomials on X . Let $A = \{p_1 \geq 0, \dots, p_m \geq 0\}$ be a set of constraints (on $X = (X_1, \dots, X_n)$). 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 $q = r_1^2 + \dots + r_\ell^2$ where $r_1, \dots, r_\ell \in \mathbb{R}[X]$. For $S \subseteq [m]$, we use p_S to denote $\prod_{i \in S} p_i$ with $p_\emptyset = 1$. Now, suppose that for all $S \subseteq [m]$ there exists $q_S \in \mathbb{M}[X]$ such that

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

Then, it is clear that the constraint set A is infeasible over \mathbb{R}^n . The surprisingly powerful theorem of Stengle [Ste74] (and earlier shown by Krivine [Kri64]) shows that whenever A is infeasible, such a certificate of infeasibility always exists. This theorem is known as Stengle’s Positivstellensatz. 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 [Put93].

While these results were well-known in the algebraic geometry community and are intimately tied to Hilbert’s seventeenth problem [Hil88], the interest in the theoretical computer science community is relatively new. The first to view Stengle’s positivstellensatz as a proof system for refutation were Grigoriev and Vorobjov [GV01] (It should be mentioned that an earlier paper [LMR96] also considered the proof theoretic aspects of Positivstellensatz but no attempt was made to quantify the complexity of such proofs). From the point of view of complexity theory, it is interesting to consider restricted proof systems where one only looks at proofs of refutation where $\max \deg(p_S \cdot q_S) \leq d$. We refer to this as the degree- d SoS hierarchy. This is essentially the dual of $d/2$ -level of the Lasserre hierarchy [Las01].

The reason to consider the degree- d SoS hierarchy 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 a fixed d), the degree- d SoS hierarchy is *effective* in the sense that 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 (see Parrillo [Par00] and Lasserre [Las01]). It should be mentioned that the so called *Lasserre hierarchy* [Las01] and the SoS hierarchy are essentially duals of each other. So, for the subsequent discussion, whenever we use the term Lasserre hierarchy, we mean the Lasserre / SoS hierarchy.

Given that the degree- d SoS hierarchy is automatizable, several researchers tried to understand the limitations of its power. Grigoriev [Gri01] 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 [Sch08].

Applications to hardness of approximation: While the results of Parrillo [Par00] and Lasserre [Las01] 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 [Par00, Las01] (see [BRS11, CS08]). 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 Lovasz and Schrijver, would suffice.

However, in a recent exciting paper, Barak *et al.* [BBH⁺12] used the degree-8 SoS hierarchy to refute the known integrality gap instances for Unique Games [KV05, RS09, KPS10]. 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 essentially the only 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” [RS09] or “Approximate Lasserre” [KS09]. 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 degree-8 SoS hierarchy refutes the known integrality gap instances for Unique Games [KV05, RS09, KPS10]. Analyzing the true optimum of these instances uses tools from analysis like hypercontractivity [Bon70, Bec75], the KKL theorem [KKL88] etc. Hence, to show that 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 a little later, 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, [BBH⁺12] prove SoS versions of results like hypercontractivity, small-set expansion etc.

Extending the results of [BBH⁺12], O’Donnell and Zhou [OZ13] analyze the problems “upward” of unique games like MAX-CUT and BALANCED-SEPARATOR. In particular, [OZ13] refutes the integrality gap instances of balanced separator from [DKSV06]. Since the key to analyzing the optimum of the BALANCED-SEPARATOR instances in [DKSV06] is the KKL theorem [KKL88], 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 [KV05]. More precisely, for any $\rho \in (-1, 0)$, [KV05] 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 [MOO10]. Thus, to refute these instances completely i.e. show that the true optimum is $\arccos \rho/\pi + o(1)$, the authors essentially needed to prove the Majority is Stablest theorem in some constant degree- d SoS hierarchy. While the authors could not prove that, they do manage to prove the weaker “ $2/\pi$ ” theorem from [KKMO07] in (some constant degree of) the SoS hierarchy. 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 [Lor86] and a powerful matrix version of Putinar’s Positivstellensatz [Las10] to prove that a certain polynomial approximation preserves positiveness. We mention here that unlike the previous two papers [BBH⁺12, OZ13] connecting SoS hierarchy with hardness of approximation, we make essential use of Putinar’s Positivstellensatz (i.e. essentially the completeness of the SoS hierarchy). The following is the main theorem concerning the power of SoS hierarchy on MAX-CUT instances.

Theorem. SoS-version of MAX-CUT *For every $\delta \in (0, 1)$ and $\rho \in (-1, 0)$, $\exists d = d(\delta, \rho)$ such that the degree- d SoS hierarchy can certify that the MAX-CUT instances from [KV05] with noise ρ have true optimum less than $\arccos \rho/\pi + \delta$.*

As the key intermediate theorem, we establish a SoS version of the well-known version of the Majority is Stablest theorem (for $\rho \in (-1, 0)$) which is stated next informally.

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 $0 \leq f(x) \leq 1$ for all $x \in \{-1, 1\}^n$ and $\max_i \text{Inf}_i(f) \leq \tau$. There is a degree- d SoS proof of the statement $\text{Stab}_\rho(f) \geq 1 - \arccos \rho/\pi - \delta - c \cdot \tau$.*

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 its difficult to express the precise statement without having the proper notation. However, we do remark that SoS version of MAX-CUT follows easily by composing the proof of refutation of UNIQUE-GAMES instances of [KV05] (done in [BBH⁺12]) along with the [KKMO07] reduction (the proof of soundness of this reduction is the step where we require the SoS version of Majority is Stablest).

2 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 $X = (x_1, \dots, x_n)$ and let $\mathbb{R}[X]$ be the ring of real polynomials over these variables. We also let $\mathbb{R}_{\leq d}[X]$ 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. We next define a set of constraints given as :

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

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

$$\mathcal{C}_d(A_e) = \{p_s(X) \cdot p(X) : s \in [m], p(X) \in \mathcal{M}_{n,d}[X] \text{ and } \deg(p) + \deg(p_s) \leq d\}$$

$$\mathcal{C}_d(A_g) = \left\{ \prod_{i=1}^m q_i^{a_i}(X) : a_1, \dots, a_m \in \mathbb{Z}^+ \text{ and } \sum_{i=1}^m a_i \cdot \deg(q_i) \leq d \right\}$$

Note that A_g includes $1 \in \mathbb{R}$.

Fact 2.1. Given A_e and A_g as described above, the sets $\mathcal{C}_d(A_e)$ and $\mathcal{C}_d(A_g)$ can be computed in time $n^{O(d)} \cdot m$.

It is obvious that without loss of generality, we can impose the constraints : $p(X) \in \mathcal{C}_d(A_e)$, $p(X) = 0$ and for $q(X) \in \mathcal{C}_d(A_g)$, $q(X) \geq 0$. Next define,

$$\mathcal{C}_d(A) = \{p(X) = 0 : \forall p(X) \in \mathcal{C}_d(A_e)\} \cup \{q(X) \geq 0 : \forall q(X) \in \mathcal{C}_d(A_g)\}$$

Definition 2.2. For the constraint set $A = A_e \cup A_g$ defined above and $h(X) \in \mathbb{R}[X]$, we say $A \vdash_d h(X) \geq 0$ 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 $q(X) \in \mathcal{C}_d(A_g)$, $\deg(r_q) + \deg(q) \leq d$. In this case, we say that A degree- d SoS proves $h(X) \geq 0$.

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 [OZ13]. The reason we are interested in Definition 2.2 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 $\tilde{E} : \mathbb{R}_{\leq d}[X] \rightarrow \mathbb{R}$ satisfying the following conditions (see [Par00] for more details)

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

A map \tilde{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 [Las01] and Parrilo [Par00] have shown that using semidefinite programming, it is possible to decide the feasibility of such a map \tilde{E} in time $m \cdot n^{O(d)}$. In fact, if there exists such a map \tilde{E} , then the algorithm outputs one in the same time. It is important to mention that since \tilde{E} has an infinite domain, it is not obvious what one means by outputting the map. To see why this makes sense, note that \tilde{E} is a linear map and hence it suffices to give to specify \tilde{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.* [BBH⁺12].

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

Fact 2.3. • 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 $A \vdash_d \{p_1 \geq 0, p_2 \geq 0, \dots, p_m \geq 0\}$ and $\{p_1 \geq 0, p_2 \geq 0, \dots, p_m \geq 0\} \vdash_{d'} q \geq 0$, $A \vdash_{d+d'} q \geq 0$.

Several other SoS facts are proven in the Appendix C. We suggest that the non-expert reader look at the Appendix C to get more comfortable with the notion of SoS proofs. 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. $X = (x_1, \dots, x_n)$) 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. To get the reader more acquainted with the power of SoS proofs, we state the following powerful result of Putinar which we use repeatedly.

Theorem 2.4. [Put93] *Let $A = \{p_1(X) \geq 0, \dots, p_m(X) \geq 0\}$ and define $\mathbb{M}(A) = \sum_{i=1}^n r_i p_i + r_0$ where $r_0, \dots, r_m \in \mathbb{M}[X]$. 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 $p \in \mathbb{M}(A)$.*

As a key step in one of our proofs, we will also require a matrix version of Putinar's Positivstellensatz (see [Las10] 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 2.5. *Let $A = \{p_1(X) \geq 0, \dots, p_m(X) \geq 0\}$ be satisfying the conditions in the hypothesis of Theorem 2.4. 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 $\Gamma_0, \dots, \Gamma_m$ are sum-of-squares.*

3 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 [MOO10] and the Gaussian stability inequality of Borell [Bor85]. We begin by defining the following function from [MN12] for every $\rho \in [-1, 1] : J_\rho : (0, 1)^2 \rightarrow [0, 1]$ as

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

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

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

Definition 3.1. *Let Ω_1 and Ω_2 be probability spaces 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 $f : \Omega_1 \rightarrow \mathbb{R}$ and $g : \Omega_2 \rightarrow \mathbb{R}$ with $\mathbf{E}_\mu f = \mathbf{E}_\mu g = 0$,*

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

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 set $Y = X$, and with probability $1 - \rho$, choose Y independently from ν . If μ is the distribution of (X, Y) , then it is easy to check that μ has Rényi correlation ρ .

Definition 3.2. *If Ω is a probability space and f is a function $\Omega^n \rightarrow \mathbb{R}$, then for $X \in \Omega$, we define $f_X : \Omega^{n-1} \rightarrow \mathbb{R}$ by*

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

Definition 3.3. *For a function $f : \Omega \rightarrow \mathbb{R}$, define*

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

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

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

noting that $\mathbf{E}[f_{X_n}|X_n]$ is a function $\Omega \rightarrow \mathbb{R}$.

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

Theorem 3.4. *For any $\epsilon > 0$ 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 $f : \Omega_1^n \rightarrow [\epsilon, 1 - \epsilon]$ and $g : \Omega_2^n \rightarrow [\epsilon, 1 - \epsilon]$,*

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

We note that [MN12] proved that in the *Gaussian setup* where $f, g : \mathbb{R}^n \rightarrow [0, 1]$ it was shown that $\mathbf{E} J_\rho(f(X), g(Y)) \leq J_\rho(\mathbf{E} f, \mathbf{E} g)$.

3.1 The base case We prove Theorem 3.4 by induction on n . In this section, we will prove the base case $n = 1$:

Claim 3.5. *For any $\epsilon > 0$ and $0 < \rho < 1$, there is a $C(\rho)$ such that for any two random variables $X, Y \in [\epsilon, 1 - \epsilon]$ with correlation in $[0, \rho]$,*

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

The proof of Claim 3.5 essentially follows from Taylor's theorem applied to the function J ; the crucial point is that J satisfies a certain differential equation. Define the matrix $M_{\rho\sigma}(x, y)$ by

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

Claim 3.6. *For any $(x, y) \in (0, 1)^2$ and $0 \leq \sigma \leq \rho$, $M_{\rho\sigma}(x, y)$ is a negative semidefinite matrix. Likewise, if $\rho \leq \sigma \leq 0$, then $M_{\rho\sigma}(x, y)$ is a positive semidefinite matrix.*

Claim 3.7. *For any $-1 < \rho < 1$, there exists $C(\rho) > 0$ such that for any $i, j \geq 0$, $i + j = 3$,*

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

Claims 3.6 and 3.7 follow from elementary calculus, and we defer their proofs to the appendix (Claim 3.6 is first proved in [MN12] and we include the proof here for the sake of completeness). Now we will use them with Taylor's theorem to prove Claim 3.5.

Proof of Claim 3.5. Fix $\epsilon > 0$ and $\rho \in (0, 1)$, and let $C(\epsilon)$ be large enough so that all third derivatives of J are uniformly bounded by $C(\epsilon)$ on the square $[\epsilon, 1 - \epsilon]^2$ (such a $C(\epsilon)$ exists by Claim 3.7). Taylor's theorem then implies that for any $a, b, a + x, b + y \in [\epsilon, 1 - \epsilon]$,

$$\begin{aligned} J_\rho(a + x, b + y) &\leq J_\rho(a, b) + x \frac{\partial J}{\partial x}(a, b) + y \frac{\partial J}{\partial y}(a, b) \\ &\quad + \frac{1}{2}(x y) \begin{pmatrix} \frac{\partial^2 J_\rho}{\partial x^2}(a, b) & \frac{\partial^2 J_\rho}{\partial x \partial y}(a, b) \\ \frac{\partial^2 J_\rho}{\partial x \partial y}(a, b) & \frac{\partial^2 J_\rho}{\partial y^2}(a, b) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + C(\rho)\epsilon^{-C(\rho)}(x^3 + y^3). \end{aligned} \quad (2)$$

Now suppose that X and Y are random variables taking values in $[\epsilon, 1 - \epsilon]$. If we apply (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

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

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 3.6.

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

Applying this to (3), we obtain

$$\mathbf{E} J_\rho(X, Y) \leq J_\rho(\mathbf{E}X, \mathbf{E}Y) + C(\rho)\epsilon^{-C(\rho)}(\mathbf{E}|\tilde{X}|^3 + \mathbf{E}|\tilde{Y}|^3)$$

□

3.2 The inductive step Next, we prove Theorem 3.4 by induction.

Proof of Theorem 3.4. Suppose that the Theorem holds with n replaced by $n - 1$. Consider $f : \Omega_1^n \rightarrow [\epsilon, 1 - \epsilon]$ and $g : \Omega_2^n \rightarrow [\epsilon, 1 - \epsilon]$.

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 ,

$$\begin{aligned} \mathbf{E}_{\tilde{X}, \tilde{Y}} J_\rho(f_{X_n}(\tilde{X}), g_{Y_n}(\tilde{Y})) \\ \leq J_\rho(\mathbf{E}[f_{X_n}|X_n], \mathbf{E}[g_{Y_n}|Y_n]) + C(\rho)\epsilon^{-C(\rho)}(\Delta_{n-1}(f_{X_n}) + \Delta_{n-1}(g_{Y_n})). \end{aligned} \quad (4)$$

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

$$\begin{aligned} \mathbf{E}_{X_n, Y_n} J_\rho(\mathbf{E}[f_{X_n}|X_n], \mathbf{E}[g_{Y_n}|Y_n]) \\ \leq J_\rho(\mathbf{E}f, \mathbf{E}g) + C(\rho)\epsilon^{-C(\rho)}(\Delta_1(\mathbf{E}[f_{X_n}|X_n]) + \Delta_1(\mathbf{E}[g_{Y_n}|Y_n])). \end{aligned} \quad (5)$$

Taking the expectation of (4) and combining it with (5), we obtain

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

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

$$J_\rho(\mathbf{E}f, \mathbf{E}g) + C(\rho)\epsilon^{-C(\rho)}(\Delta_n(f) + \Delta_n(g)).$$

□

4 Borell's inequality

The most interesting special case of Theorem 3.4 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. First, let us recall the functional version of Borell's inequality that was given in [MN12].

Theorem 4.1. Suppose that 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 $f_1, f_2 : \mathbb{R}^d \rightarrow [0, 1]$,

$$\mathbf{E} J(f_1(G_1), f_2(G_2)) \leq J(\mathbf{E} f_1, \mathbf{E} f_2).$$

We will prove Theorem 4.1 using Theorem 3.4 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 4.2. For $X \in \{-1, 1\}^n$, define

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

Then

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

Proof. The proof is by induction: the base case is trivial, while the inductive step follows by Jensen’s inequality:

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

□

Proof of Theorem 4.1. Let $n = md$ and, for each $i = 1, \dots, d$, define

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

Define $G_{2,n}$ similarly by with Y instead of X . By the multivariate central limit theorem, $(G_{1,n}, G_{2,n}) \xrightarrow{d} (G_1, G_2)$ as $m \rightarrow \infty$.

Suppose first that f_1 and f_2 are L -Lipschitz functions taking values in $[\epsilon, 1 - \epsilon]$, and define $g_1, g_2 : \{-1, 1\}^n$ by $g_i(X) = f_i(G_{1,n})$. By Theorem 3.4,

$$\mathbf{E} J(g_1(X), g_2(Y)) \leq J(\mathbf{E} g_1, \mathbf{E} g_2) + C(\epsilon)(\Delta_n(g_1) + \Delta_n(g_2)). \quad (6)$$

Since f_i is L -Lipschitz,

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

for every j , and so Claim 4.2 implies that

$$\Delta_n(g_i) \leq \frac{8L^3 n}{m^{3/2}} = \frac{8L^3 d}{\sqrt{m}}.$$

Applying this to (6),

$$\mathbf{E} J(g_1(X), g_2(Y)) \leq J(\mathbf{E} g_1, \mathbf{E} g_2) + C(\epsilon) \frac{16L^3 d C(\epsilon)}{\sqrt{m}}$$

and so the definition of g_i implies

$$\mathbf{E} J(f_1(G_{1,n}), f_2(G_{2,n})) \leq J(\mathbf{E} f_1(G_{1,n}), \mathbf{E} f_2(G_{2,n})) + C(\epsilon) \frac{16L^3 dC(\epsilon)}{\sqrt{m}}$$

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

$$\mathbf{E} J(f_1(G_1), f_2(G_2)) \leq J(\mathbf{E} f_1(G_1), \mathbf{E} f_2(G_2)). \quad (7)$$

This establishes the theorem for functions f_1 and f_2 which are Lipschitz and take values in $[\epsilon, 1 - \epsilon]$. But any measurable $f_1, f_2 : \mathbb{R}^d \rightarrow [0, 1]$ can be approximated (say in $L^p(\mathbb{R}^d, \gamma_d)$) by Lipschitz functions with values in $[\epsilon, 1 - \epsilon]$. Since neither the Lipschitz constant nor ϵ appears in (7), the general statement of the theorem follows from the dominated convergence theorem. \square

5 Majority is stablest

By giving a bound on $\Delta_n(f)$ that is better than Claim 4.2, we can derive the ‘‘Majority is Stablest’’ theorem from Theorem 3.4. 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 max influence of f . For this, we will introduce some very basic Fourier analytic preliminaries below.

Fourier analysis: We start by defining the ‘‘character’’ functions i.e. for every $S \subseteq [n]$, define $\chi_S(x) : \{-1, 1\}^n \rightarrow \mathbb{R}$ as $\chi_S(x) = \prod_{i \in S} x_i$. Now, every function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ can be expressed as

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

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 $\sum_{S \subseteq [n]} \widehat{f}^2(S) = \mathbf{E}_{x \in \{-1, 1\}^n} [f^2(x)]$. This is known in literature as Parseval’s identity. 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 $f : \{-1, 1\}^n \rightarrow \mathbb{R}$, $T_\rho f(x) = \mathbf{E}_{y \sim_\rho x} [f(y)]$. 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 [O’D07] 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 [Mos05] 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 3.2:

Definition 5.1. For disjoint sets $S, T \subseteq [n]$, 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 $f : \{-1, 1\}^n \rightarrow \mathbb{R}$, a set $S \subseteq [n]$, and an element $x \in \{-1, 1\}^S$, we define $f_x : \{-1, 1\}^{[n] \setminus S} \rightarrow \mathbb{R}$ 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 5.2. If $S_i = \{i + 1, \dots, n\}$, 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 $f : \{-1, 1\} \rightarrow \mathbb{R}$,

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

For the inductive step, we have

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

□

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

Claim 5.3. 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 $g : \{-1, 1\}^n \rightarrow \mathbb{R}$ 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

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

□

In particular, the identity of Claim 5.3 allows us to compute second moments of $\widehat{f_X}$:

Claim 5.4. For any function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$, any $x \in \{-1, 1\}^S$ and any $i \in U \subset [n]$,

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

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

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

Proof. In view of Claim 5.3, we can write

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

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

$$\mathbf{E}_{X_S} |\widehat{f_{X_S}}(U)|^2 = \sum_{T \subset S} \widehat{f}^2(T \cup U). \quad (8)$$

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

For the second part,

$$\sum_{i=1}^n \mathbf{E}_{X \in \{-1,1\}^{S_i}} |\widehat{f}_X(i)|^2 = \sum_{i=1}^n \sum_{T \subset S_i} \widehat{f}^2(T \cup \{i\}) = \sum_{U \subset [n], U \neq \emptyset} \widehat{f}^2(U), \quad (9)$$

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 (9) is just $\text{Var}(f)$. \square

Next, we will consider $\widehat{f}_x(n-i)$ as a polynomial in x and apply hypercontractivity to the right hand side of Claim 5.2. First, note that T_σ commutes (up to a multiplicative factor) with restriction:

Claim 5.5. *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 5.3,

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

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

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

\square

Essentially, Claim 5.5 allows us to apply the Bonami-Beckner inequality to \widehat{f}_X : for any $\sigma < 1$, if $p = 1 + \sigma^{-2}$ then

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

By Claim 5.4, if $i \in U$ then

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

Applying this to $S_i = \{i+1, \dots, n\}$ and $U_i = \{i\}$ and summing the result over $i = 1, \dots, n$, we obtain

$$\sum_i \mathbf{E}_{X \in \{-1,1\}^{S_i}} |(\widehat{T_\sigma f})_X(i)|^p \leq (\max_i \text{Inf}_i(f))^{p/2} \text{Var}(f).$$

Now, if f takes values in $[-1, 1]$ then $\text{Var}(f) \leq 1$ and all Fourier coefficients of f (and its restriction) are bounded by 1. Hence, Claim 5.2 implies the following:

Claim 5.6. *If $1 + \sigma^{-2} \leq 3$ then*

$$\Delta_n(T_\sigma f) \leq (\max_i \text{Inf}_i(f))^{1-\sigma^2/2}.$$

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

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

As remarked earlier, our proof extends to the generalizations of Theorem 5.7 such as those presented by [DMR06, Mos10]. 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 [Mos10] for more details.

Proof. Suppose $f : \{-1, 1\}^n \rightarrow [\epsilon, 1 - \epsilon]$ satisfies $\max_i \text{Inf}_i(f) \leq \tau$, and let X, Y be uniformly random elements of $\{-1, 1\}^n$ with $\mathbf{E} X_i Y_i = \rho$. For sufficiently small $\eta > 0$, Claim 5.6 implies that

$$\Delta_n(T_{1-\eta}f) \leq \tau^{c\eta}.$$

Note that the range of $T_{1-\eta}f$ belongs to $[\epsilon, 1 - \epsilon]$ because the range of f does. Hence, Theorem 3.4 applied to $T_{1-\eta}f$ implies that

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

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

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

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

$$\mathbb{S}_\rho(f) = \mathbf{E} T_{1-\eta}f(X)T_{1-\eta}f(Y) \leq J_{\rho(1-\eta)^2}(\mathbf{E} f, \mathbf{E} f) + C\epsilon^{-C(\rho)}\tau^{c\eta} \quad (10)$$

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

Now, (10) holds for any $f : \{-1, 1\}^n \rightarrow [\epsilon, 1 - \epsilon]$. For a function f taking values in $[-1, 1]$, let f^ϵ be f truncated to $[\epsilon, 1 - \epsilon]$. Since $|\mathbf{E} f^\epsilon - \mathbf{E} f| \leq \epsilon$ and (by the proof of Claim 3.6) $\frac{\partial J_\rho(x, y)}{\partial x} \leq 1$ for any ρ ,

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

On the other hand, $|f - f^\epsilon| \leq \epsilon$ and so

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

Thus, (10) applied to f^ϵ implies that for any $\rho \leq (1-\eta)^2$ and any $\epsilon > 0$,

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

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

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

Finally, some calculus on J_ρ (see Claim A.1) shows that $|\frac{\partial J_\rho(x, y)}{\partial \rho}| \leq (\sqrt{1-\rho^2})^{-3/2}$ for any x, y ; hence

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

Choosing $\eta = C(\rho) \frac{\log \log(1/\tau)}{\log(1/\tau)}$ completes the proof as long as $\rho \leq (1-\eta)^2$. However, we can trivially make the theorem true for $(1-\eta)^2 \leq \rho$ by choosing $C(\rho)$ and $c(\rho)$ appropriately. \square

6 SoS proof of Majority is Stablest

The principal theorem of this section is the SoS version of ‘‘Majority is Stablest’’ theorem of [MOO10]. 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

$$A_\rho = \{0 \leq f(x) \leq 1 : \text{for all } x \in \{-1, 1\}^n\}$$

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

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

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.

$$\text{Inf}_i(f) = \sum_{S \ni i} \widehat{f}^2(S) \quad \text{Inf}_i^{\leq d}(f) = \sum_{S \ni i: |S| \leq d} \widehat{f}^2(S)$$

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*

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

This is easily seen to be equal to the statement of the Majority is Stablest theorem of [MOO10]. Before, we delve further into the SoS proofs, we feel its good to familiarize ourselves with the fourier machinery in the SoS world. The upshot of the discussion ahead is going to be that the basic fourier identities and operations hold without any changes in the SoS world. In particular, it is easy to verify 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 $g(x) = \mathbf{E}_{y \sim_{\rho} x} [f(y)]$ 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 Smoothing the function For our purposes, it is necessary to do a certain smoothing of the function f . In particular, we start by considering a new function f_2 i.e. we create a new sequence of indeterminates defined by $f_2(x) = (1 - \epsilon)f(x) + \epsilon/2$ for some $\epsilon > 0$. The value of ϵ shall be fixed later. We observe that

$$A_p \vdash_1 \cup_{x \in \{-1,1\}^n} \{\epsilon \leq f_1(x) \leq 1 - \epsilon\} \quad (11)$$

$$\widehat{f}_1(S) = (1 - \epsilon)\widehat{f}(S) + (\epsilon/2) \cdot \mathbf{1}_{S=\emptyset}$$

We also make the following claim (the proof is deferred to Appendix D).

Claim 6.2.

$$A_p \vdash_2 f(x)f(y) - 2\epsilon \leq f_1(x)f_1(y) \leq f(x)f(y) + 2\epsilon$$

The next stage of smoothing is done by defining $g = T_{1-\eta}f_2$ for some $\eta > 0$. Again, the value of η will be fixed later.

$$\cup_{x \in \{-1,1\}^n} \{\epsilon \leq f_1(x) \leq 1 - \epsilon\} \vdash_1 \cup_{x \in \{-1,1\}^n} \{\epsilon \leq g(x) \leq 1 - \epsilon\} \quad (12)$$

$$\widehat{g}(S) = (1 - \epsilon)(1 - \eta)^{|S|} \widehat{f}(S) + (\epsilon/2) \cdot \mathbf{1}_{S=\emptyset}$$

Also, observe that $\mathbf{E}_{x,y \sim_{\rho} x} [f_2(x) \cdot f_2(y)] = \mathbf{E}_{x,y \sim_{\rho'} x} [g(x) \cdot g(y)]$ where $\rho' = \rho/(1 - \eta)^2$. Of course, this imposes the condition $|\rho| < |1 - \eta|^2$. So, we have to choose η to be small enough. Now, define the constraint set $A'_p = \cup_{x \in \{-1,1\}^n} \{\epsilon \leq g(x) \leq 1 - \epsilon\}$. 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 $A'_p \vdash_d q \geq 0$, then $A_p \vdash_d q \geq 0$.*

The proof of the above is obtained by combining (11) and (12) with the third bullet of Fact 2.3. The second claim is

Claim 6.4.

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

Thus, the above two claims mean that from now on, we will work with A'_p and aim to prove a lower bound on $\mathbb{E}_{x,y \sim \rho', x} [g(x) \cdot g(y)]$. At this stage, let $\tilde{J}_{\rho'}$ be the approximation obtained from Claim B.1 with parameter $\epsilon > 0$ and $\delta = \epsilon$. 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 $\epsilon > 0$, such that $\rho' \in (-1, 0)$ and \tilde{J} is as described above, there is a $d_\alpha = d_\alpha(\epsilon, \rho')$ such that,*

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

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 $(x, y) \in [\epsilon, 1 - \epsilon]^2$, $xy \geq \tilde{J}(x, y) - \epsilon$. In other words, if we define the polynomial $p(x, y) = xy - \tilde{J}(x, y) + 2\epsilon$, then we know that for $(x, y) \in [\epsilon, 1 - \epsilon]^2$, $p(x, y) \geq \epsilon$. We can thus apply Corollary C.10 to get that there is an integer $d_\alpha = d_\alpha(\epsilon, \rho')$ such that for $(x, y) \in [\epsilon, 1 - \epsilon]^2$, for $\rho' \in (0, 1)$,

$$\{\epsilon \leq x \leq 1 - \epsilon, \epsilon \leq y \leq 1 - \epsilon\} \vdash_{d_1} p \geq 0$$

Expanding p , finishes the proof. \square

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 3.5.

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 $A = \cup_{i,j \in \{0,1\}} \{\epsilon \leq h_i(j) \leq 1 - \epsilon\}$. For any $\epsilon > 0$, $\rho' \in (-1, 0)$, $\exists c_\gamma = c_\gamma(\epsilon, \rho')$ and $\exists d_\gamma = d_\gamma(\epsilon, \rho')$ such that*

$$A \vdash_{d_\gamma} \mathbb{E}_{\substack{x \in \mathbb{R}^{\{-1,1\}} \\ y \sim \rho', x}} [\tilde{J}(h_0(x), h_1(y))] \geq \tilde{J}(\widehat{h}_0(0), \widehat{h}_1(0)) - \epsilon \cdot (\widehat{h}_0^2(1) + \widehat{h}_1^2(1)) - c_\gamma \cdot (\widehat{h}_0^4(1) + \widehat{h}_1^4(1))$$

where $\widehat{h}_i(j) = \frac{h_i(0) + (-1)^j \cdot h_i(1)}{2}$ for $i, j \in \{0, 1\}$.

Proof. We start by noting that since \tilde{J} is a symmetric polynomial, hence we can write

$$\tilde{J}(x, y) = \sum_{m,n:m+n \leq K} \mu_{\{m,n\}} x^m y^n$$

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

Claim 6.7.

$$\mathbb{E}_{\substack{x \in \mathbb{R}^{\{-1,1\}} \\ y \sim \rho', x}} [\tilde{J}(h_0(x), h_1(y))] = \sum_{m,n:m+n \text{ is even}} \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \cdot \left(\frac{1 + \rho'}{2} + (-1)^m \cdot \frac{1 - \rho'}{2} \right)$$

where

$$\nu_{m,n} = \sum_{m_1 \geq m; n_1 \geq n} \mu_{m_1, n_1} \cdot \widehat{h}_0^{m_1 - m}(0) \cdot \widehat{h}_1^{n_1 - n}(0) \cdot \binom{m_1}{m} \binom{n_1}{n}$$

The proof is deferred to Appendix D. Next, we note that $\nu_{0,0} = \tilde{J}(\widehat{h}_0(0), \widehat{h}_1(0))$. Thus, we get that

$$\mathbb{E}_{\substack{x \in \mathbb{R}^{\{-1,1\}} \\ y \sim \rho', x}} [\tilde{J}(h_0(x), h_1(y))] - \tilde{J}(\widehat{h}_0(0), \widehat{h}_1(0)) = \sum_{\substack{m,n:m+n \text{ is even} \\ K \geq m+n \geq 2}} \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \cdot \left(\frac{1 + \rho'}{2} + (-1)^m \cdot \frac{1 - \rho'}{2} \right) \quad (13)$$

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

Claim 6.8.

$$A \vdash_{4K+3} Y \geq \sum_{\substack{m,n:m+n \text{ is even} \\ \text{and } m+n \geq 4}} \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \cdot \left(\frac{1+\rho'}{2} + (-1)^m \cdot \frac{1-\rho'}{2} \right) \geq -Y$$

where $Y = 2cK^4 2^{2K} (\widehat{h}_0^4(1) + \widehat{h}_1^4(1))$.

Again, we defer the proof of Claim 6.8 to Appendix D. Thus, we are only left with the task of controlling the terms when $m+n=2$. Note that

$$\sum_{m+n=2} \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \cdot \left(\frac{1+\rho'}{2} + (-1)^m \cdot \frac{1-\rho'}{2} \right) = \nu_{2,0} \cdot \widehat{h}_0^2(1) + \nu_{0,2} \cdot \widehat{h}_1^2(1) + \rho' \nu_{1,1} \cdot \widehat{h}_0(1) \widehat{h}_1(1)$$

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

$$\frac{\partial^2 \widetilde{J}(x, y)}{\partial x^2} \Big|_{x=\widehat{h}_0(0), y=\widehat{h}_1(0)} = \frac{\partial^2 \widetilde{J}(\widehat{h}_0(0) + x', \widehat{h}_1(0) + y')}{\partial x'^2} \Big|_{x'=0, y'=0}$$

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

$$\Lambda = \frac{1}{2} \left(\frac{\partial^2 \widetilde{J}(x, y)}{\partial x^2} \widehat{h}_0^2(1) + \frac{\partial^2 \widetilde{J}(x, y)}{\partial y^2} \widehat{h}_1^2(1) + 2\rho' \frac{\partial^2 \widetilde{J}(x, y)}{\partial x \partial y} \widehat{h}_0(1) \widehat{h}_1(1) \right)$$

In the above, all the derivatives are evaluated at $x = \widehat{h}_0(0), y = \widehat{h}_1(0)$. We now make the following claim which gives a lower bound on Λ .

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

$$A \vdash_{d'_\gamma} \Lambda \geq -\epsilon \cdot (\widehat{h}_0^2(1) + \widehat{h}_1^2(1))$$

We defer this proof to Appendix D. Now, set $d_\gamma = \max\{d'_\gamma, 4K+3\}$ and $c_\gamma = 2cK^4 2^{2K}$. Combining Claim 6.8 and Claim 6.9 with (13), we get Lemma 6.6. \square

6.3 Tensorization: We now do a ‘‘tensorization’’ of the inequality in Lemma 6.6. 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: $\phi_y(z) = \phi(z \cdot y)$. 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 homogenous linear forms in the indeterminates $\widehat{\phi}(S)$ (for $S \subseteq [n]$). We now state a few basic properties for the indeterminates $g_y(z)$ and $\widehat{g}_y(S)$.

$$A'_p \vdash_1 \bigcup_{i=0}^{n-1} \bigcup_{y \in \{-1, 1\}^i} \bigcup_{z \in \{-1, 1\}^{n-i}} \{\epsilon \leq g_y(z) \leq 1 - \epsilon\} \quad (14)$$

$$A'_p \vdash_1 \bigcup_{i=0}^{n-1} \bigcup_{y \in \{-1, 1\}^i} \bigcup_{S \subseteq [n-i]} \{-1 \leq \widehat{g}_y(S) \leq 1\} \quad (15)$$

$$\vdash_2 \mathbf{E}_{y \in \{-1, 1\}^i} [\widehat{g}_x^2(n-i)] = \sum_{\substack{S \subseteq \{n-i, \dots, n\} \\ n-i \in S}} \widehat{g}^2(S) \quad (16)$$

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

$$\begin{aligned} A'_p \vdash_{d_\gamma} \mathbf{E}_{\substack{x \in \{-1,1\}^n \\ y \sim_{\rho'} x}} [\tilde{J}(g(x), g(y))] &\geq \tilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) - \epsilon \left(\sum_{i=0}^{n-1} \mathbf{E}_{z \in \{-1,1\}^i} [\hat{g}_z^2(n-i)] \right) \\ &\quad - c_\gamma \left(\sum_{i=0}^{n-1} \mathbf{E}_{z \in \{-1,1\}^i} [\hat{g}_z^4(n-i)] \right) \end{aligned}$$

The proof of this claim is a very simple induction. For the sake of completeness, we give the proof in Appendix D. We now simplify the error terms. Towards this, note that (16) implies that

$$\vdash_2 \epsilon \left(\sum_{i=0}^{n-1} \mathbf{E}_{z \in \{-1,1\}^i} [\hat{g}_z^2(n-i)] \right) = \sum_{S \neq \emptyset} \hat{g}^2(S) \leq \sum_S \hat{g}^2(S) = \mathbf{E}_{x \in \{-1,1\}^n} [g^2(x)]$$

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

$$A'_p \vdash_{d_\gamma} \mathbf{E}_{\substack{x \in \{-1,1\}^n \\ y \sim_{\rho'} x}} [\tilde{J}(g(x), g(y))] \geq \tilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) - \epsilon - c_\gamma \left(\sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1,1\}^i} [\hat{g}_x^4(n-i)] \right) \quad (17)$$

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). Now, this very naively gives an error term of the form $\mathbf{E}_x \hat{h}_x^4(n-i)$ and $\mathbf{E}_x \hat{\ell}_x^4(n-i)$. The latter can be easily bound using hypercontractivity. However, there does not seem to be obvious way to bound the former. This is in spite of the fact that $\mathbf{E}_x \hat{h}_x^2(n-i)$ 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 :

$$h(x) = \sum_{|S| > d_\eta} \hat{g}(S) \chi_S(x) \quad \ell(x) = \sum_{|S| \leq d_\eta} \hat{g}(S) \chi_S(x)$$

By the way it is defined, it is clear that $\vdash_1 h(x) + \ell(x) = g(x)$. Now, we can analyze the term $\mathbf{E}_{x \in \{-1,1\}^i} [\hat{g}_x^4(n-i)]$ as

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

We begin by stating the following useful fact :

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

Proof.

$$\vdash_2 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1,1\}^i} \hat{h}_x^2(n-i) = \sum_S \hat{h}^2(S) = \sum_{|S| > d_\eta} (1-\eta)^{d_\eta} \hat{f}^2(S) \leq \eta \cdot \left(\sum_{|S| > d_\eta} \hat{f}^2(S) \right) \leq \eta \cdot \left(\sum_{|S|} \hat{f}^2(S) \right)$$

$$A_p \vdash_3 \left(\sum_{|S|} \hat{f}^2(S) \right) = \mathbf{E}_{x \in \{-1,1\}^n} [f^2(x)] \leq 1$$

Combining the two facts, finishes the proof. \square

We now make the following claims.

Claim 6.12. $A_p \vdash_6 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1,1\}^i} [\widehat{g}_x^3(n-i) \widehat{h}_x(n-i)] \leq \sqrt{\eta}$.

Claim 6.13. $A_p \vdash_5 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1,1\}^i} [\widehat{g}_x^2(n-i) \widehat{\ell}_x^2(n-i)] \leq \sqrt{\eta} + \frac{9^{d_\eta}}{\sqrt{\eta}} (\sum_{i=1}^n (\text{Inf}_i^{\leq d_\eta}(f))^2)$

Claim 6.14. $A_p \vdash_8 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1,1\}^i} [\widehat{g}_x^2(n-i) \widehat{h}_x(n-i) \widehat{\ell}_x(n-i)] \leq 2\sqrt{\eta} + \frac{9^{d_\eta}}{\sqrt{\eta}} (\sum_{i=1}^n (\text{Inf}_i^{\leq d_\eta}(f))^2)$

The proofs are deferred to the appendix. Combining (17) with Claim 6.12, Claim 6.13, Claim 6.14 along (17) and Claim 6.3, we get that for c_γ and d_γ described in Lemma 6.6,

$$\begin{aligned} A_p \vdash_{d_\gamma} \mathbf{E}_{\substack{x \in \{-1,1\}^n \\ y \sim_{\rho'} x}} [\widetilde{J}(g(x), g(y))] &\geq \widetilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) - \epsilon - 4 \cdot c_\gamma \sqrt{\eta} \\ &- \frac{2 \cdot 9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}} \left(\sum_{i=1}^n (\text{Inf}_i^{\leq d_\eta}(f))^2 \right) \end{aligned}$$

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

$$A_p \vdash_{d_\alpha} \mathbf{E}_{x, y \sim_{\rho'} x} [f(x) \cdot f(y)] \geq \mathbf{E}_{x, y \sim_{\rho'} x} [\widetilde{J}(g(x), g(y))] - 4\epsilon \quad (18)$$

Thus, now applying (18), we get

$$A_p \vdash_{\max\{d_\gamma, d_\alpha\}} \mathbf{E}_{\substack{x \in \{-1,1\}^n \\ y \sim_{\rho'} x}} [f(x) \cdot f(y)] \geq \widetilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) - 5\epsilon - 4 \cdot c_\gamma \sqrt{\eta} - \frac{2 \cdot 9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}} \left(\sum_{i=1}^n (\text{Inf}_i^{\leq d_\eta}(f))^2 \right) \quad (19)$$

Now, define a new sequence of indeterminates $\{f_c(x)\}_{x \in \{-1,1\}^n}$ where $f_c(x) = 1 - f(x)$. Next, define $f_{2c}(x) = (1 - \epsilon)f_c(x) + \epsilon/2$. Next, we define $g_c(x) = \mathbf{E}_{y \sim_{1-\eta} x} [f_{2c}(x)]$. We now make the following observations :

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

Thus, using the above, analogous to (19), we have the following :

$$A_p \vdash_{\max\{d_\gamma, d_\alpha\}} \mathbf{E}_{\substack{x \in \{-1,1\}^n \\ y \sim_{\rho'} x}} [f_c(x) \cdot f_c(y)] \geq \widetilde{J}(\mathbf{E}[g_c(x)], \mathbf{E}[g_c(y)]) - 5\epsilon - 4 \cdot c_\gamma \sqrt{\eta} - \frac{2 \cdot 9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}} \left(\sum_{i=1}^n (\text{Inf}_i^{\leq d_\eta}(f_c))^2 \right) \quad (20)$$

Now, define ξ as $\xi = 5\epsilon - 4 \cdot c_\gamma \sqrt{\eta} - \frac{2 \cdot 9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}} (\sum_{i=1}^n (\text{Inf}_i^{\leq d_\eta}(f))^2)$. Summing up (19) and (20), we get

$$\begin{aligned} A_p \vdash_{\max\{d_\gamma, d_\alpha\}} \mathbf{E}_{\substack{x \in \{-1,1\}^n \\ y \sim_{\rho'} x}} [f(x) \cdot f(y) + (1 - f(x)) \cdot (1 - f(y))] \\ \geq \widetilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) + \widetilde{J}(\mathbf{E}[1 - g(x)], \mathbf{E}[1 - g(y)]) - 2\xi \end{aligned} \quad (21)$$

Next, we recall the following fact :

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

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

Combining this fact with Claim B.1, we have that for every $x \in [\epsilon, 1 - \epsilon]$,

$$\tilde{J}(x, x) + \tilde{J}(1 - x, 1 - x) \geq 1 - \frac{\arccos \rho'}{\pi} - 2\epsilon$$

By using Corollary C.10, we have that there exists $d_\delta = d_\delta(\epsilon, \rho')$ such that

$$A_p \vdash_{d_\delta} \tilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) + \tilde{J}(\mathbf{E}[1 - g(x)], \mathbf{E}[1 - g(y)]) \geq 1 - \frac{\arccos \rho'}{\pi} - 4\epsilon \quad (22)$$

Combining (21) and (22), we get

$$A_p \vdash_{\max\{d_\gamma, d_\alpha, d_\delta\}} \mathbf{E}_{\substack{x \in \{-1, 1\}^n \\ y \sim_\rho x}} [f(x) \cdot f(y) + (1 - f(x)) \cdot (1 - f(y))] \geq 1 - \frac{\arccos \rho'}{\pi} - 14\epsilon - 8 \cdot c_\gamma \sqrt{\eta} - \frac{4 \cdot 9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}} \left(\sum_{i=1}^n (\text{Inf}_i^{\leq d_\eta}(f))^2 \right) \quad (23)$$

From here, getting to Theorem 6.1 is pretty easy. We proceed as follows :

- For the given ρ and κ , first we choose $\epsilon = \kappa/100$. This implies that $14\epsilon \leq \kappa/4$.
- Next, observe that $c_\gamma(\rho', \epsilon)$ is a uniformly continuous function of ρ' and ϵ . Now, recall that $\rho' = \rho/(1 - \eta)$. Hence, there exists $\eta_0 = \eta_0(\rho, \epsilon, \kappa)$ such that for all $\eta \leq \eta_0$, $\sqrt{\eta} \cdot c_\gamma(\rho', \epsilon) \leq \kappa/32$.
- Again, observe that for any $\rho \in (-1, 0)$ and $\kappa > 0$, there exists $\eta_1 = \eta_1(\rho, \kappa)$ 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 $d_0(\kappa, \rho) = \max\{d_\gamma, d_\alpha, d_\delta\}$, $c(\kappa, \rho) = \frac{4 \cdot 9^{d_\eta} \cdot c_\gamma}{\sqrt{\eta}}$ and $d_1(\kappa, \rho) = d_\eta$ and hence get

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

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 [KV05] for the noise parameter ρ . Let $\{x_v\}_{v \in V}$ be a sequence of indeterminates and $A = \cup_{v \in V} \{0 \leq x_v \leq 1\}$. Then, for any $\delta > 0$, there exists $d_1 = d_1(\delta, \rho)$ such that*

$$A \cup \left\{ \mathbf{E}_{(u,v) \in E} x_u \cdot (1 - x_v) + x_v \cdot (1 - x_u) \geq \frac{1}{\pi} \arccos \rho + \delta \right\} \vdash_{d_1} -1 \geq 0$$

For this section, it is helpful to begin by recalling the following theorem of O'Donnell and Zhou [OZ13].

Theorem 7.2. [OZ13] *Let $\{f(x)\}_{x \in \{-1, 1\}^n}$ be a sequence of indeterminates and let $A = \cup_{x \in \{-1, 1\}^n} \{0 \leq f(x) \leq 1\}$. Then, for any $\delta > 0$ and $\rho \in (-1, 0)$,*

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

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

The main application of Theorem 7.2 in [OZ13] is the following : Khot and Vishnoi [KV05] construct instances of MAX-CUT (parameterized by noise parameter ρ) whose optimum is bounded by $(\arccos \rho)/\pi + o(1)$ (and yet the basic SDP relaxation for MAX-CUT has value $(1 - \rho)/2$.) O’Donnell and Zhou essentially use Theorem 7.2 as a black-box to give a constant degree SoS proof that these instances have optimum bounded by $1 - K(\rho) + o(1)$. This improves significantly on the bound provided by the basic SDP.

In this section, we will show how we can use the stronger version of Theorem 7.2, namely Theorem 6.1 to do even better. In particular, we will use this theorem to give a constant degree SoS proof that these MAX-CUT instances have optimum bounded by $(\arccos \rho)/\pi + o(1)$ (which is of course tight). We will not give all the details of this proof as our proof will follow the (by now, standard) reduction from [KKMO07] and its SoS variant from [OZ13].

We begin by recalling the description of instances of UNIQUE-GAMES (UG). A UG instance is specified by a set of vertices V and an alphabet $[k]$. Along with this, there is a probability distribution \mathcal{E} on tuples of the form $(u, v, \pi_{(u,v)})$ with $\pi_{(u,v)} : [k] \rightarrow [k]$ being a permutation. Further, the weighted graph defined by \mathcal{E} is regular. Also, let \mathcal{E}_u denote the marginal distribution on (v, π) when the first vertex is conditioned to be u . The objective is to get a mapping $L : V \rightarrow [k]$ so as to maximize the following quantity :

$$\Pr_{(u,v,\pi_{(u,v)}) \in \mathcal{E}}[\mathcal{L}(v) = \pi_{(u,v)}(\mathcal{L}(u))]$$

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 [OZ13]. In particular, we define variables $x_{v,i}$ for every $v \in V$ and $i \in [k]$. Now, consider the set of constraints defined by

$$A_p = \bigcup_{v \in V, i \in [k]} \{x_{v,i} \geq 0\} \bigcup_{v \in V} \left\{ \sum_{i \in [k]} x_{v,i} \leq 1 \right\}$$

It is easy to show that if the optimum solution to the Unique Games instance is bounded by β , then

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

for any set of indeterminates $\{x_{v,i}\}$ which obeys the constraint set A_p . We now make the following definition :

Definition 7.3. *Given a UG instance (V, \mathcal{E}) with alphabet size k , there is a degree- d SOS refutation for optimum β if*

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

Before we go ahead, we recall that for any $\eta \in (0, 1)$ and $N \in \mathbb{N}$ (which is a power of 2), [KV05] construct UG instances over $2^N/N$ vertices, alphabet size n such that optimal value of the instance is bounded by $N^{-\eta}$.¹ Modifying the result from [BBH⁺12], O’Donnell and Zhou [OZ13] show the following :

Theorem 7.4. *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 [KV05]. Then, there is a degree-4 SoS refutation for optimum $\beta = N^{-\Omega(\eta)}$.*

We next describe the reduction from [KKMO07] 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 E_u which is defined as the marginal of E conditioned on the first vertex being u .
- Choose $x \in \{-1, 1\}^k$ and $y \sim_{\rho} x$.

¹Of course, the interesting part is that [KV05] shows that the standard SDP relaxation on this instance has value $1 - \eta$

- Output vertices $((v_1, \pi_{(u,v_1)}(x)), (v_2, \pi_{(u,v_2)}(y)))$

Now, for a function $g : \{-1, 1\}^k \rightarrow [0, 1]$, let us define $\text{Stab}_\rho(g)$ as follows

$$\text{Stab}_\rho(g) = \mathbf{E}_{x \in \{-1, 1\}^k, y \sim_\rho x} [g(x) \cdot g(y) + (1 - g(x)) \cdot (1 - g(y))]$$

We have the following simple claim :

Claim 7.5. [KKMO07] *Let $G' = (V', E_{\rho,k})$ be an instance of MAX-CUT described above. Consider a partition of the graph G' (into two sets) specified by a collection of functions $\{f_v : \{-1, 1\}^k \rightarrow \{0, 1\}\}$. Then, the value of cut defined by this partition is $1 - \mathbf{E}_{u \in V} [\text{Stab}_\rho(g_u)]$ where*

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

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 $A_m = \cup_{v \in V} \cup_{z \in \{-1, 1\}^k} \{0 \leq f_v(z) \leq 1\}$. Then, O'Donnell and Zhou [OZ13] show that if (V, \mathcal{E}) is a UG instance such that there is a degree 4 refutation for the optimum β , then

$$A_m \cup \{1 - \mathbf{E}_{u \in V} [\text{Stab}_\rho(g_u)] \geq K(\rho) - \delta - 2^{O(1/\delta^2)} \beta\} \vdash_{O(1/\delta^2)+4} -1 \geq 0 \quad (24)$$

This of course means that the if the [KKMO07] reduction is applied on the instances from Theorem 7.4,

$$A_m \cup \{1 - \mathbf{E}_{u \in V} [\text{Stab}_\rho(g_u)] \geq K(\rho) - \delta - 2^{O(1/\delta^2)} \cdot N^{-\Omega(\eta)}\} \vdash_{O(1/\delta^2)+4} -1 \geq 0$$

Exactly following the same steps as [OZ13], but using Theorem 6.1 instead of Theorem 7.2, we show that for any $\delta > 0$,

$$A_m \cup \{1 - \mathbf{E}_{u \in V} [\text{Stab}_\rho(g_u)] \geq (\arccos \rho)/\pi - \delta - d_2(\delta, \rho)c(\delta, \rho) \cdot \beta\} \vdash_{d_1(\delta, \rho)+4} -1 \geq 0 \quad (25)$$

We do not repeat the steps here and leave it to the reader to fill the details. Using $\beta = N^{-\Omega(\eta)}$, we get Theorem 7.1.

Acknowledgements

We thank Ryan O'Donnell and Yuan Zhou for sharing the manuscript [OZ13].

References

- [Aus07] P. Austrin. Balanced MAX-2-SAT might not be the hardest. In *Proc. 39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 189--197. ACM, 2007.
- [BBH⁺12] B. Barak, F. G. S. L. Brandão, A. Harrow, J. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proc. 44th Annual ACM Symposium on Theory of Computing (STOC)*, pages 307--326, 2012.
- [Bec75] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102:159--182, 1975.
- [Bob97] S. 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.
- [Bon70] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(fasc. 2):335--402 (1971), 1970.
- [Bor85] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Probability Theory and Related fields*, 70:1--13, 1985.

- [Bou02] J. Bourgain. On the distributions of the fourier spectrum of boolean functions. *Israel J. Math.*, 131:269--276, 2002.
- [BRS11] B. Barak, P. Raghavendra, and D. Steurer. Rounding Semidefinite Programming Hierarchies via Global Correlation. In *Proc. 52nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 472--481, 2011.
- [CS08] E. Chlamtac and G. Singh. Improved Approximation Guarantees through Higher Levels of SDP Hierarchies. In *APPROX-RANDOM*, pages 49--62, 2008.
- [DKSV06] N. Devanur, S. Khot, R. Saket, and N. Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *Proc. 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 537--546, 2006.
- [DMR06] I. Dinur, E. Mossel, and O. Regev. Conditional hardness for approximate coloring. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing (STOC 2006)*, pages 344--353, 2006.
- [DS05] I. Dinur and S. Safra. On the hardness of approximating minimum vertex-cover. *Annals of Mathematics*, 162(1):439--485, 2005.
- [FK96] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124:2993--3002, 1996.
- [FKN08] E. Friedgut, G. Kalai, and N. Nisan. Elections Can be Manipulated Often. In *Proc. 49th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 243--249, 2008.
- [FM98] D. Felsenthal and M. Machover. *The Measurement of Voting Power*. Edward Elgar Publishing, 1998.
- [Gri01] D. Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613--622, 2001.
- [GV01] D. Grigoriev and N. Vorobjov. Complexity of Null-and Positivstellensatz proofs. *Ann. Pure Appl. Logic*, 113(1-3):153--160, 2001.
- [GW95] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Jour. of the ACM*, pages 1115--1145, 1995.
- [Hås97] J. Håstad. Some optimal inapproximability results. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1--10, May 1997.
- [Hil88] D. Hilbert. Über die Darstellung definiter Formen als Summe von Formenquadraten. *Mathematische Annalen*, 12:307--326, 1888.
- [IM12] M. Isaksson and E. Mossel. New maximally stable gaussian partitions with discrete applications. *Israel Journal of Mathematics*, 189:347--396, 2012.
- [Kal02] G. Kalai. A Fourier-theoretic perspective on the Condorcet paradox and Arrow's theorem. *Advances in Applied Mathematics*, 29(3):412--426, 2002.
- [Kal04] G. Kalai. Social Indeterminacy. *Econometrica*, 72:1565--1581, 2004.
- [Kho02] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th Ann. STOC*, pages 767--775, 2002.
- [KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proc. 29th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 68--80, 1988.
- [KKMO07] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319--357, 2007.
- [KO12] G. Kindler and R. O'Donnell. Gaussian noise sensitivity and fourier tails. In *IEEE Conference on Computational Complexity*, pages 137--147, 2012.

- [KPS10] S. Khot, P. Popat, and R. Saket. Approximate Lasserre Integrality Gap for Unique Games. In *APPROX-RANDOM*, pages 298--311, 2010.
- [Kri64] J. Krivine. Anneaux préordonnés. *J. Analyse Math.*, 12:307--326, 1964.
- [KS09] S. Khot and R. Saket. SDP Integrality Gaps with Local ℓ_1 -Embeddability. In *Proc. 50th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 565--574, 2009.
- [KV05] S. Khot and N. K. Vishnoi. The Unique Games Conjecture, Integrality Gap for Cut Problems and Embeddability of Negative Type Metrics into l_1 . In *Proc. 46th Symposium on Foundations of Computer Science (FOCS)*, pages 53--62, 2005.
- [Las01] J. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796--817, 2001.
- [Las10] J. Lasserre. *Moments, Positive polynomials and their Applications*. Imperial College Press, London, 2010.
- [LMR96] H. Lombardi, N. Mnev, and M-F. Roy. The Positivstellensatz and Small Deduction Rules for Systems of Inequalities. *Mathematische Nachrichten*, 181(1):245--259--, 1996.
- [Lor86] G.G. Lorentz. *Bernstein polynomials*. Chelsea Publishing Company, Incorporated, 1986.
- [MN12] E. Mossel and J. Neeman. Robust optimality of gaussian noise stability. 2012.
- [MOO10] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171:295--341, 2010.
- [Mos05] E. Mossel. Lecture notes on Fourier Analysis. <http://www.stat.berkeley.edu/~mossel/teach/206af05/>, 2005.
- [Mos10] E. Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713--1756, 2010.
- [Mos12] E. Mossel. A quantitative Arrow theorem. *Probab. Theory Related Fields*, 154(1--2):49--88, 2012.
- [O'D07] R. O'Donnell. Analysis of boolean functions. <http://www.cs.cmu.edu/odonnell/boolean-analysis/>, 2007.
- [OW08] R. O'Donnell and Y. Wu. An optimal SDP algorithm for Max-Cut, and equally optimal long code tests. In *Proc. 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 335--344, 2008.
- [OZ13] R. O'Donnell and Y. Zhou. Approximability and Proof Complexity. In *Proc. 24th Annual Symposium on Discrete Algorithms (SODA)*, 2013.
- [Par00] P. Parrilo. *Structured semidefinite programs and Semialgebraic Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000.
- [Put93] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969--984, 1993.
- [Rag08] P. Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proc. 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 245--254, 2008.
- [RS09] P. Raghavendra and D. Steurer. Integrality Gaps for Strong SDP Relaxations of UNIQUE GAMES. In *Proc. 50th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 575--585, 2009.
- [Sch08] G. Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k-CSPs. In *Proc. 49th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 593--602, 2008.
- [She99] W. Sheppard. On the application of the theory of error to cases of normal distribution and normal correlation. *Phil. Trans. Royal Soc. London*, 192:101--168, 1899.

- [ST06] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influence of variables, and PCPs. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 11--20, New York, NY, USA, 2006. ACM Press.
- [Ste74] G. Stengle. A Nullstellensatz and a Positivstellensatz in Semialgebraic Geometry. *Mathematische Annalen*, 207(2):87--97, 1974.
- [Tal94] M. Talagrand. On Russo's approximate 0-1 law. *Ann. Probab.*, 22(3):1576--1587, 1994.

A Facts regarding J_ρ

Here we collect various facts about the function

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

where $(X, Y) \sim \mathcal{N}(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$. These calculations all follow from elementary calculus.

Claim 3.6. For any $(x, y) \in (0, 1)^2$ and $0 \leq \sigma \leq \rho$, $M_{\rho\sigma}(x, y)$ is a negative semidefinite matrix. Likewise, if $\rho \leq \sigma \leq 0$, 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

$$K_\rho(s, t) = \Pr_{X, Y}[X \leq s, Y \leq t] = \Pr_{X, Z}[X \leq s, Z \leq (t - \rho \cdot X) / \sqrt{1 - \rho^2}]$$

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

$$K_\rho(s, t) = \int_{s'=-\infty}^s \phi(s') \int_{t'=-\infty}^{(t-\rho \cdot s')/\sqrt{1-\rho^2}} \phi(t') ds' dt' \quad (26)$$

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

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

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 $|\sigma| \leq |\rho|$. Then

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

If $\rho \geq 0$ 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 $\rho \leq 0$ then the diagonal is non-negative and so $M_{\rho\sigma}(x, y)$ is positive semidefinite. \square

Claim 3.7. For any $-1 < \rho < 1$, there exists $C(\rho) > 0$ such that for any $i, j \geq 0$, $i + j = 3$,

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

Proof. As before, we set $\Phi^{-1}(x) = s$ and $\Phi^{-1}(y) = t$. From the proof of Claim 3.6, 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 $\frac{\partial s}{\partial x} = \frac{1}{\phi(s)}$ and $\frac{\partial t}{\partial y} = \frac{1}{\phi(t)}$, we have

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

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

$$\begin{aligned} \exp \left(- \frac{t^2 - 2\rho st + (3\rho^2 - 2)s^2}{2(1 - \rho^2)} \right) &= e^{-\frac{t^2}{2(1-\rho^2)}} e^{\frac{\rho st}{1-\rho^2}} e^{\frac{(2-3\rho^2)s^2}{2(1-\rho^2)}} \\ &\leq e^{-\frac{t^2}{2(1-\rho^2)}} e^{\frac{\rho(s^2+t^2)}{2(1-\rho^2)}} e^{\frac{(2-3\rho^2)s^2}{2(1-\rho^2)}} \\ &\leq (x(1-x)y(1-y))^{-\frac{\rho}{2(1-\rho^2)}} (x(1-x))^{-\frac{2-3\rho^2}{2(1-\rho^2)}} \\ &\leq (x(1-x)y(1-y))^{-C(\rho)}. \end{aligned}$$

Applying this to (27), we see that

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

The other third derivatives are similar:

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

By the same steps that led to (27), we get

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

(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 . \square

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

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

Proof. We begin from (26), 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 $\phi \leq 1$ and $\int_{s'} \phi(s') ds' = 1$, it follows that

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

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. \square

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 $0 \leq i + j \leq 2$,

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

Moreover, if $\rho \in [-1 + \epsilon, 1 - \epsilon]$, then the degree of \tilde{J} and the maximal coefficient in \tilde{J} can be bounded by constants depending only on ϵ and δ .

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 [Lor86].

Theorem B.2. Suppose $f : [0, 1] \rightarrow \mathbb{R}$ 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 $0 \leq i \leq m$,

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

Seeing as the first three derivatives of J_ρ are bounded on $[\epsilon, 1 - \epsilon]$, 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 $f : [0, 1]^2 \rightarrow \mathbb{R}$ 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 $0 \leq i + j \leq 2$ and note that

$$\frac{\partial^j g_n(\cdot, y)}{\partial y^j} = B_n \frac{\partial^j f(\cdot, y)}{\partial y^j} \tag{27}$$

$$\frac{\partial^i h_n(x, \cdot)}{\partial x^i} = B_n \frac{\partial^i g_n(x, \cdot)}{\partial x^i}. \tag{28}$$

Now, fix $y \in [0, 1]$ and apply Theorem B.2 to (27): for any $x \in [0, 1]$,

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

On the other hand, fixing x and applying Theorem B.2 to (28) yields

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

Putting these together,

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

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 $\rho \in [-1 + \epsilon, 1 - \epsilon]$ since the third derivatives of J_ρ are uniformly bounded over $\rho \in [-1 + \epsilon, 1 - \epsilon]$. Moreover, since $\max_{x,y} |J_\rho(x, y)| \leq 1$, the coefficients in h_n can be bounded in terms of n , which is in turn bounded in terms of ϵ and δ . \square

C Useful facts in SoS hierarchy

Fact C.1. *If $A \vdash_d p \geq 0$ and $A \vdash q \geq 0$, then $A \vdash_d p + q \geq 0$.*

Fact C.2. *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 [OZ13])

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

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 $A \vdash_k -1 \leq y^k$. \square

Fact C.3. $-1 \leq y \leq 1 \vdash_5 y^4 \leq y^2$

Proof.

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

This finishes the proof. \square

Fact C.4. *Let $a \leq y \leq b \vdash_d p(y) \geq 0$. Then, for $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ such that $\forall i \in [k], \lambda_i \geq 0$ and $\sum_{i=1}^k \lambda_i = 1$, we have that $\{\cup_{i=1}^k 0 \leq Z_i \leq 1\} \vdash_d p(\sum_{i=1}^k \lambda_i Z_i) \geq 0$.*

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 $\{\cup_{i=1}^k 0 \leq Z_i \leq 1\} \vdash_d p(\sum_{i=1}^k \lambda_i Z_i) \geq 0$. \square

Fact C.5. For $k \geq 3$, $-1 \leq y \leq 1 \vdash_{2k+1} 0 \leq y^{2k} \leq y^4$

Proof. $1 - y^2 = \frac{1}{2}(1+y)^2(1-y) + \frac{1}{2}(1+y)(1-y)^2$ 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. \square

Fact C.6. 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 C.5, $-1 \leq y \leq 1 \vdash_{2m+1} y^4 \geq y^{2m}$. Similarly, we have $-1 \leq z \leq 1 \vdash_{2m+1} z^4 \geq z^{2m}$. 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. \square

Fact C.7. 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 first use Fact 3.10 from [OZ13] which states that $A \vdash_4 yz^3 \leq y^4 + z^4$. We can replace y by $-y$ to get the other inequality. This already gives the proof for $n = 3$. For $n > 3$, we have that $A \vdash_{n+1} yz^n \leq y^4 z^{n-3} + z^{n+1}$. Now, using Fact C.2 (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. \square

Fact C.8. Let $A \vdash_{d_1} 0 \leq x \leq 1$ and $A \vdash_{d_2} -z \leq y \leq z$ where $z \in \mathbb{M}_{\leq d}[X]$. 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. \square

Fact C.9. [BBH⁺12] 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,

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

Consequences of Putinar's Positivstellensatz The following is a consequence of Theorem 2.4.

Corollary C.10. Let $X = (x_1, x_2)$ and $A = \{x_1 \geq \epsilon, x_2 \geq \epsilon, x_1 \leq 1 - \epsilon, x_2 \leq 1 - \epsilon\}$. Then, for any $p(X)$ such that $p(X) \geq \epsilon$ on $\mathbb{V}(A)$, there exists an integer $d = d(p)$ such that $A \vdash_d p \geq \epsilon/2$.

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

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

Clearly, $q(x_1, x_2) \in \mathbb{M}(S)$ 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 2.4 to get that there is an integer $d = d(p)$ such that $S \vdash_d p - \epsilon/2 \geq 0$. This implies $S \vdash p \geq \epsilon/2$. \square

The following is a corollary of Theorem 2.5.

Corollary C.11. Let $X = (x_1, \dots, x_n)$ and $A = \{p_1(X) \geq 0, \dots, p_m(X) \geq 0\}$ be satisfying the conditions in Theorem 2.4. 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 2.5, 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

D Missing proofs from Section 6

Proof (of Claim 6.2):

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

Hence, $A_p \vdash_2 f_1(x)f_1(y) - f(x)f(y) \leq 2\epsilon$. On the other hand, note that $1 - f(x)f(y) = (1 - f(x))f(y) + (1 - f(y))$ and hence $A_p \vdash_2 f(x)f(y) \leq 1$. This implies that $A_p \vdash_2 f(x)f(y) - f_1(x)f_1(y) \leq 2\epsilon$. \square

Proof (of Claim 6.7): We begin by observing that

$$\tilde{J}(\widehat{h}_0(0) + x, \widehat{h}_1(0) + y) = \sum_{m,n} \mu_{\{m,n\}} \cdot (\widehat{h}_0(0) + x)^m \cdot (\widehat{h}_1(0) + y)^n = \sum_{m,n} \nu_{m,n} \cdot x^m \cdot y^n$$

As a consequence, we get that

$$\begin{aligned} \tilde{J}(\widehat{h}_0(0) + \widehat{h}_0(1), \widehat{h}_1(0) + \widehat{h}_1(1)) &= \sum_{m,n} \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \\ \tilde{J}(\widehat{h}_0(0) - \widehat{h}_0(1), \widehat{h}_1(0) - \widehat{h}_1(1)) &= \sum_{m,n} (-1)^{m+n} \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \end{aligned}$$

Adding these equations, we get

$$\tilde{J}(\widehat{h}_0(0) + \widehat{h}_0(1), \widehat{h}_1(0) + \widehat{h}_1(1)) + \tilde{J}(\widehat{h}_0(0) - \widehat{h}_0(1), \widehat{h}_1(0) - \widehat{h}_1(1)) = 2 \cdot \sum_{\substack{m,n \\ m+n \text{ is even}}} \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \quad (29)$$

Similarly, we have that

$$\begin{aligned} \tilde{J}(\widehat{h}_0(0) - \widehat{h}_0(1), \widehat{h}_1(0) + \widehat{h}_1(1)) &= \sum_{m,n} (-1)^m \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \\ \tilde{J}(\widehat{h}_0(0) + \widehat{h}_0(1), \widehat{h}_1(0) - \widehat{h}_1(1)) &= \sum_{m,n} (-1)^n \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \end{aligned}$$

Thus,

$$\tilde{J}(\widehat{h}_0(0) - \widehat{h}_0(1), \widehat{h}_1(0) + \widehat{h}_1(1)) + \tilde{J}(\widehat{h}_0(0) + \widehat{h}_0(1), \widehat{h}_1(0) - \widehat{h}_1(1)) = 2 \cdot \sum_{\substack{m,n \\ m+n \text{ is even}}} (-1)^m \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \quad (30)$$

Hence, combining (29) and (30), we get that

$$\mathbf{E}_{\substack{x \in_R \{-1,1\} \\ y \sim_{\rho'} x}} [\tilde{J}(h_0(x), h_1(y))] = \sum_{\substack{m,n \\ m+n \text{ is even}}} \nu_{m,n} \cdot \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \cdot \left(\frac{1 + \rho'}{2} + (-1)^m \cdot \frac{1 - \rho'}{2} \right)$$

\square

Proof (of Claim 6.8): 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 2.3, 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 consistently throughout this proof. Applying Fact C.2, we get that

$$A \vdash_{m_1-m+1} 0 \leq \widehat{h}_0^{m_1-m}(0) \leq 1 \quad A \vdash_{n_1-n+1} 0 \leq \widehat{h}_1^{n_1-n}(0) \leq 1$$

and hence we have that

$$A \vdash_{m_1+n_1-m-n+2} 0 \leq \widehat{h}_0^{m_1-m}(0) \cdot \widehat{h}_1^{n_1-n}(0) \leq 1 \quad (31)$$

Now, we consider two possibilities : Either $m, n \geq 2$ or $\max\{m, n\} \geq 3$. In the first case, using Fact C.6, we get that

$$A \vdash_{1+\max\{2m, 2n\}} -(\widehat{h}_0^4(1) + \widehat{h}_1^4(1)) \leq \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \leq \widehat{h}_0^4(1) + \widehat{h}_1^4(1) \quad (32)$$

Next, consider the other case i.e. when $\max\{m, n\} \geq 3$. Without loss of generality, assume $m \geq 3$ and $n = 1$. Then, by Fact C.7, we get that

$$A \vdash_{m+2} -(\widehat{h}_0^4(1) + \widehat{h}_1^4(1)) \leq \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \leq \widehat{h}_0^4(1) + \widehat{h}_1^4(1)$$

Now, combining (31) along with an application of Fact C.8, we get that

$$A \vdash_{m+n+m_1+n_1+3} -(\widehat{h}_0^4(1) + \widehat{h}_1^4(1)) \leq \widehat{h}_0^m(1) \cdot \widehat{h}_1^n(1) \cdot \widehat{h}_0^{m_1-m}(0) \cdot \widehat{h}_1^{n_1-n}(0) \leq \widehat{h}_0^4(1) + \widehat{h}_1^4(1)$$

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

$$A \vdash_{4K+3} -c2^{2K}(\widehat{h}_0^4(1) + \widehat{h}_1^4(1)) \leq \Gamma_{m_1, n_1, m, n} \leq c2^{2K}(\widehat{h}_0^4(1) + \widehat{h}_1^4(1))$$

As,

$$\sum_{\substack{m, n: m+n \text{ is even} \\ \text{and } m+n \geq 4}} \nu_{m,n} \widehat{h}_0^m(1) \widehat{h}_1^n(1) \cdot \left(\frac{1+\rho'}{2} + (-1)^m \cdot \frac{1-\rho'}{2} \right) = \sum_{\substack{m, n: m+n \text{ is even} \\ m+n \geq 4 \\ K \geq m_1 \geq m \quad K \geq n_1 \geq n}} \Gamma_{m_1, n_1, m, n}$$

As a result, we can conclude that

$$A \vdash_{4K+3} -c2^{2K} K^4 (\widehat{h}_0^4(1) + \widehat{h}_1^4(1)) \leq \sum_{\substack{m, n: m+n \text{ is even} \\ \text{and } m+n \geq 4}} \nu_{m,n} \widehat{h}_0^m(1) \widehat{h}_1^n(1) \leq c2^{2K} K^4 (\widehat{h}_0^4(1) + \widehat{h}_1^4(1))$$

□

Proof (of Claim 6.9): 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 $\beta = 2\epsilon$. Now, let us define $\widetilde{M}_1 = \widetilde{M} + \beta I$. Using Claim 3.6 and Claim B.1, for $\rho' \in (-1, 0)$, and $(x, y) \in [\epsilon, 1 - \epsilon]^2$, we can say that $\widetilde{M}_1 \succeq \epsilon \cdot I$. Hence, using Corollary C.11, $\exists d'_\gamma$ 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 the degree d'_γ depends only on ϵ and ρ is because from Corollary C.11, the degree d'_γ depends on ϵ , ρ and the polynomial \tilde{J} which again in turn depends only on ϵ and ρ . □

Proof (of Lemma 6.10): The proof is by induction. We introduce the following notation : For any $z \in \{-1, 1\}^i$, we use $1 \cdot z \in \{-1, 1\}^{i+1}$ to denote the string z with a 1 prefixed to it. Likewise, we define $-1 \cdot z \in \{-1, 1\}^{i+1}$ to denote the string z with a -1 prefixed to it. Next, for any pairs of strings $z_{-1}, z_1 \in \{-1, 1\}^i$ and $j, k \in \{-1, 1\}$, we define the indeterminate, $h_i(j) = \mathbf{E}_{z \in \{-1, 1\}^{n-i-1}} [g(z \cdot j \cdot z_i)]$. Define $A_{z_{-1}, z_1} = \cup_{j, k \in \{-1, 1\}} \{\epsilon \leq h_i(j) \leq 1 - \epsilon\}$. 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

$$\begin{aligned} A_{z_{-1}, z_1} \vdash_{d_\gamma} & \mathbf{E}_{\substack{x \in \{-1, 1\} \\ y \sim_{\rho'} x}} \tilde{J}(\mathbf{E}[g_{x \cdot z_1}], \mathbf{E}[g_{y \cdot z_{-1}}]) \geq \tilde{J}(\mathbf{E}[g_{z_1}], \mathbf{E}[g_{z_{-1}}]) - \epsilon \left(\widehat{g}_{z_1}^2(n-i) + \widehat{g}_{z_{-1}}^2(n-i) \right) \\ & - c_\gamma \left(\widehat{g}_{z_{-1}}^4(n-i) + \widehat{g}_{z_1}^4(n-i) \right) \end{aligned}$$

As a consequence of the third bullet of Fact 2.3, 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. We multiply the inequality by $(\frac{1+\rho'}{4})^{i-d(z_1, z_{-1})} \cdot (\frac{1-\rho'}{4})^{d(z_1, z_{-1})}$. Note that we will consider the case when $i = 0$ and hence z_1 and z_{-1} is the empty string. In this scenario, $d(z_1, z_{-1})$ is defined to be zero. We now consider all the above inequalities generated by choosing $(z_1, z_{-1}) \in \{-1, 1\}^i \times \{-1, 1\}^i$ for $0 \leq i < n$. Next, we add all these inequalities (Fact C.1) but the degree of the resulting SoS proof remains d_γ . Now, it is easy to see that all terms of the form $\tilde{J}(\mathbf{E}[g_{z_1}], \mathbf{E}[g_{z_{-1}}])$ cancel out except when $z_1, z_{-1} \in \{-1, 1\}^n$ or $z_1 = z_{-1} = \phi$.

$$\mathbf{E}_{\substack{x \in \{-1, 1\}^n \\ y \sim_{\rho'} x}} [\tilde{J}(g(x), g(y))] \geq \tilde{J}(\mathbf{E}[g(x)], \mathbf{E}[g(y)]) + \text{error terms}$$

We now compute the error terms. First, we sum up the error coming from the term $\epsilon \left(\widehat{g}_{z_1}^2(n-i) + \widehat{g}_{z_{-1}}^2(n-i) \right)$. For any given $z_1 \in \{-1, 1\}^i$, consider the term $\beta \widehat{g}_{z_1}^2(n-i)$. For every $z_{-1} \in \{-1, 1\}^i$, it occurs with the factor $\frac{1+\rho'}{4}^{i-d(z_1, z_{-1})} \cdot \frac{1-\rho'}{4}^{d(z_1, z_{-1})}$. 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 $\epsilon \left(\widehat{g}_{z_1}^2(n-i) + \widehat{g}_{z_{-1}}^2(n-i) \right)$. The calculation of the ‘fourth degree’ error terms is exactly identical resulting in the final theorem. \square

Proof (of Claim 6.12): We have

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

Next, recall that using Fact 6.11, we have $A_p \vdash_3 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1, 1\}^i} \widehat{h}_x^2(n-i) \leq \eta$. Similarly, from (15) and Claim 6.3, we have that $A_p \vdash -1 \leq \widehat{g}_x(n-i) \leq 1$. This in turn implies $A_p \vdash_7 \widehat{g}_x^6(n-i) \leq \widehat{g}_x^2(n-i)$ (combining Fact C.3 and Fact C.5)

Combining all the above, we get

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

However, again we have that $A_p \vdash_3 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1, 1\}^i} \widehat{g}_x^2(n-i) = \sum_{|S|>0} \widehat{g}^2(S) \leq \mathbf{E}[g^2(x)] \leq 1$. This gives us the claim. \square

Proof (of Claim 6.13): We have

$$\begin{aligned} \vdash_4 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1, 1\}^i} [\widehat{g}_x^2(n-i) \widehat{\ell}_x^2(n-i)] & \leq \frac{\sqrt{\eta}}{2} \left(\sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1, 1\}^i} \widehat{g}_x^4(n-i) \right) + \frac{\left(\sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1, 1\}^i} \widehat{\ell}_x^4(n-i) \right)}{2\sqrt{\eta}} \\ \vdash_4 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1, 1\}^i} \widehat{\ell}_x^4(n-i) & \leq 9^{d_\eta} \cdot \left(\sum_{i=0}^{n-1} \left(\mathbf{E}_{x \in \{-1, 1\}^i} \widehat{\ell}_x^2(n-i) \right)^2 \right) \quad (\text{using Fact C.9}) \end{aligned}$$

As in the proof of Claim 6.12, we can show $A_p \vdash_5 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1,1\}^i} \widehat{g}_x^4(n-i) \leq 1$. Combining all the above, we get

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

Again, observe that

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

By using the second bullet of Fact 2.3, we can also get

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

Combining these, we get the final result. \square

Proof (of Claim 6.14):

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

From the proof of Claim 6.12, we know that $A_p \vdash_3 \sum_{i=0}^{n-1} \mathbf{E}_{x \in \{-1,1\}^i} \widehat{h}_x^2(n-i) \leq \eta$. Thus, we get

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

However,

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

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

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

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

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

Combining these, we have the proof. \square