Electronic Colloquium on Computational Complexity, Report No. 75 (2013)

A Characterization of Strong Approximation Resistance

Subhash Khot* Madhur Tulsiani† Pratik Worah‡

May 23, 2013

Abstract

For a predicate $f : \{-1,1\}^k \mapsto \{0,1\}$ with $\rho(f) = \frac{|f^{-1}(1)|}{2^k}$, we call the predicate *strongly approximation resistant* if given a near-satisfiable instance of CSP(f), it is computationally *hard* to find an assignment such that the fraction of constraints satisfied is outside the range $[\rho(f) - \Omega(1), \rho(f) + \Omega(1)]$.

We present a characterization of strongly approximation resistant predicates under the Unique Games Conjecture. We also present characterizations in the *mixed* linear and semi-definite programming hierarchy and the Sherali-Adams linear programming hierarchy. In the former case, the characterization coincides with the one based on UGC. Each of the two characterizations is in terms of existence of a probability measure on a natural convex polytope associated with the predicate.

The predicate is called *approximation resistant* if given a near-satisfiable instance of CSP(f), it is computationally *hard* to find an assignment such that the fraction of constraints satisfied is at least $\rho(f)+\Omega(1)$. When the predicate is odd, i.e. $f(-z) = 1 - f(z), \forall z \in \{-1,1\}^k$, it is easily observed that the notion of approximation resistance coincides with that of strong approximation resistance. Hence for odd predicates, in all the above settings, our characterization of strong approximation resistance is also a characterization of approximation resistance.

*NYU and University of Chicago. khot@cims.nyu.edu Research partly supported by NSF Expeditions grant CCF-0832795, NSF Waterman Award and BSF grant 2008059.

†Toyota Technological Institute at Chicago. madhurt@ttic.edu.

‡University of Chicago. pworah@cs.uchicago.edu.

ISSN 1433-8092

1 Introduction

Constraint satisfaction problems (CSPs) are some of the most well-studied NP-hard problems. Given a predicate $f : \{-1,1\}^k \mapsto \{0,1\}$, an instance of CSP($f$) consists of n $\{-1,1\}$-valued[1] variables and m constraints where each constraint is the predicate f applied to an ordered subset of k variables, possibly in negated form. For example, the OR predicate on k variables corresponds to the k-SAT problem whereas the PARITY predicate (i.e. whether the product of the variables is $+1$) on k variables corresponds to the k-LIN problem. The satisfiability problem for CSP(f) asks whether there is an assignment that satisfies *all* the constraints. A well-known *dichotomy* result of Schaefer [34] shows that for every predicate f, the satisfiability problem for CSP(f) is either in P or NP-complete and moreover his characterization explicitly gives a (short) list of predicates for which the problem is in P.

An instance of CSP(f) is called α-satisfiable if there is an assignment that satisfies at least α fraction of the constraints. The focus of this paper is whether given a $(1 - o(1))$-satisfiable instance, there is an efficient algorithm with a *non-trivial* performance.

The density of the predicate $\rho(f) = \frac{|f^{-1}(1)|}{2^k}$ is the probability that a uniformly random assignment to its variables satisfies the predicate. Given an instance of CSP(f) with m constraints, a naive algorithm that assigns random $\{-1,1\}$ values to its variables yields an assignment such that the fraction of constraints satisfied is $\rho(f)$ in expectation and *with high probability* is in the range $[\rho(f) - o(1), \rho(f) + o(1)]$ if the instance is *reasonable* (e.g. if every variable appears in at most $o(m)$ constraints).

With this observation in mind, we consider two notions of a *non-trivial* algorithm, the first one being new (though implicit in prior literature especially from hardness side) and the second one being standard. For both, the instance is promised to be $(1-o(1))$-satisfiable. In the first notion, an algorithm is considered non-trivial if it finds an assignment such that the fraction of assignments satisfied is outside the range $[\rho(f) - \Omega(1), \rho(f) + \Omega(1)]$, i.e. the algorithm has to do something more clever than outputting a random assignment. If such an efficient algorithm exists, we call the predicate *weakly approximable* and *strongly approximation resistant* otherwise. In the second notion, one that is more well-studied, an algorithm is considered non-trivial if it finds an assignment that satisfies at least $\rho(f)+\Omega(1)$ fraction of the constraints. If such an efficient algorithm exists, the predicate is called *approximable* and *approximation resistant* otherwise. Note that an approximable predicate is also weakly approximable and as a contra-positive, a strongly approximation resistant predicate is also approximation resistant. Also, it is easily observed that for an odd predicate, i.e. $f(z) = 1 - f(z) \forall z \in \{-1,1\}^k$, the two notions are equivalent. For an odd predicate, the non-constant part of its Fourier representation has only odd degree monomials and the constant term (as always) is $\rho(f)$. Flipping the sign of all variables simultaneously if necessary, a weak approximation (i.e. *deviating* from $\rho(f)$) is easily turned into a standard approximation (i.e. *exceeding* $\rho(f)$). At the risk of jumping ahead a little, we also mention here that all prior works showing approximation resistance of specific predicates, with possibly one exception, in fact show strong approximation resistance either implicitly or explicitly.

Towards the study of resistance, it is convenient to define the gap versions of the problem. GapCSP$(f)_{c,s}$ is a promise problem such that the instance is guaranteed to be either c-satisfiable or at most s-satisfiable. Thus a predicate is approximation resistant if GapCSP$(f)_{1-o(1),\ \rho(f)+o(1)}$

[1] It is more convenient to work with $\{-1,1\}$-valued variables than $\{0,1\}$-valued ones. Here -1 corresponds to logical TRUE and $+1$ to logical FALSE. Multiplication of variables in the {-1,1}-world corresponds to XOR-ing them in the {0,1}-world.

is not in P. We define GapCSP$(f)_{1-o(1),\ \rho(f)\pm o(1)}$ as the promise problem such that the instance is guaranteed to be either $(1-o(1))$-satisfiable or for *every* assignment, the fraction of constraints satisfied is in the range $[\rho(f)-o(1), \rho(f)+o(1)]$. Thus a predicate is strongly approximation resistant if GapCSP$(f)_{1-o(1),\ \rho(f)\pm o(1)}$ is not in P. For resistant predicates, one would ideally like to show that the corresponding gap problem is NP-hard, or as is often the case, settle for a weaker notion of hardness such as UG-hardness (i.e. NP-hard assuming the Unique Games Conjecture [24]) or hardness, a.k.a. *integrality gap*, for a specific family of linear or semidefinite programming relaxation. We now give an overview of prior works, which concern only with the notion of approximation resistance, though as we mentioned, the notion of strong approximation resistance is implicit in the hardness works.

Until early 1990s, very little, if anything, was known regarding whether any *interesting* predicate is approximable or approximation resistant. By now we have a much better understanding of this issue thanks to a sequence of spectacular results. Goemans and Williamson [17, 36] showed that 2SAT and 2LIN are approximable.[2] The discovery of the PCP Theorem [15, 2, 1], aided by works such as [7, 32], eventually led to Håstad's result that 3SAT and 3LIN are approximation resistant and in fact that the appropriate gap versions are NP-hard! Since then, many predicates have been shown to be approximation resistant (see e.g. [18, 33, 23, 14], all NP-hardness) and most recently, a remarkable result of Chan [9] shows the approximation resistance of the Hypergraph Linearity Predicate (he shows NP-hardness whereas UG-hardness was shown earlier in [33]). Also, a general result of Raghavendra [29] shows that if a predicate is approximable, then it is so via a natural SDP relaxation of the problem followed by a *rounding* of the solution (the result is more general than stated: it applies to every (c,s)-gap).

In this paper, our focus is towards obtaining a complete characterization of (strong) approximation resistance for all predicates, in the spirit of Schaefer's theorem. There has been some progress in this direction that we sketch now. Every predicate of arity 2 is approximable as follows from Goemans and Williamson's algorithm [17].[3] A complete classification of predicates of arity 3 is known [36, 38]: a predicate of arity 3 is approximation resistant (NP-hard) if it is implied by PARITY up to variable negations and approximable otherwise. For predicates of arity 4, Hast [19] gives a partial classification. Austrin and Mossel [6] show that a predicate is approximation resistant (UG-hard) if the set $f^{-1}(1)$ of its satisfying assignments supports a pairwise independent distribution (for a somewhat more general sufficient condition see [4]). Using this sufficient condition, Austrin and Håstad [3] show that a vast majority of k-ary predicates for large k are approximation resistant. Hast [20] shows that a k-ary predicate with at most $k-1$ satisfying assignment is approximable.

In spite of all these works, a complete characterization of approximation resistance remains elusive. A recent result of Austrin and Khot [5] gives a complete characterization of approximation resistance (UGC-based) when the CSP is restricted to be k-partite[4] and the predicate is even.[5] Given an even predicate f, the authors therein associate with it a convex polytope $\mathcal{C}(f)$ consisting of all vectors of dimension $\binom{k}{2}$ that arise as the *second moment vectors* $(\mathbb{E}_{z\sim\nu}[z_i z_j] \,|\, 1 \le i < j \le k)$ of distributions ν supported on $f^{-1}(1)$. It is shown that the k-partite version of CSP(f) is approximation resistant (UG-hard) if and only if $\mathcal{C}(f)$ supports a distribution (a measure to be more

[2] The result is more famously known for the MAX-CUT problem, but MAX-CUT is not a CSP in our sense of the definition as it does not allow variable negations. Once variable negations are allowed, MAX-CUT is same as 2LIN.

[3] Håstad [21] shows the same for 2-ary predicates over larger alphabet as well. We restrict to boolean alphabet in this paper.

[4] Meaning the set of variables is partitioned into k layers and for every constraint, the i^{th} variable is from the i^{th} layer.

[5] Meaning $f(-z) = f(z)\ \forall z \in \{-1,1\}^k$.

precise) with a certain (difficult to state) property. The k-partiteness condition is rather restrictive and without the evenness condition, one would need to take into account the *first moment vector* $(\mathbb{E}_{z\sim\nu}[z_i]\,|1 \leq i \leq k)$ as well and it is not clear how to incorporate this in [5].

Characterizing Strong Approximation Resistance

In this paper, we make a significant (in our opinion) progress on the question of characterizing (strong) approximation resistance. Our main result is a complete characterization of strong approximation resistance. As we noted, for odd predicates, this is same as characterizing approximation resistance: one gets strong approximation resistance on the hardness side and standard approximation on the algorithmic side, i.e. best of both the worlds. One interesting family of odd predicates is balanced linear threshold functions.

Before stating the characterization, let us first point out that the characterization is not as *simple* as one may wish and we do not know whether it is decidable, both these features also shared by the result in [5]. [6] Therein the authors also argue why a *simple* characterization might be unlikely and we share that view as well. Also, while we get only a characterization for strong approximation resistance and not for the more standard notion of approximation resistance, it should be kept in mind that almost all known approximation resistance results actually show strong resistance, either implicitly or explicitly, or by a minor modification or possibly switching from NP-hardness to UG-hardness. This is typically because the soundness analysis of these constructions shows that Fourier terms that are potentially responsible for *deviating* from the threshold $\rho(f)$ are all bounded by $o(1)$ in *magnitude*, thus showing that it is hard to not just *exceed* $\rho(f)$, but even to *deviate* from it. The only possible exception we are aware of is an example of an arity 4 predicate in [4], Example 8.7 therein. The authors show that the predicate is approximation resistance by presenting a *hard to round* point (this concept is elaborated later). The same point is not good enough to show strong approximation resistance, but it is possible that there is another one or a *probability measure* on points which is good enough (we have not investigated this possibility yet).

Roughly speaking our characterization states that a predicate $f : \{-1,1\}^k \mapsto \{0,1\}$ is approximation resistant (UG-hard) if and only if a convex polytope $\mathcal{C}(f)$ associated with it supports a probability measure with certain symmetry properties.[7] Specifically, let $\mathcal{C}(f)$ be the convex polytope consisting of all vectors of dimension $k + \binom{k}{2}$ that arise as the *first and second moment vectors*

$$\left(\left(\mathop{\mathbb{E}}_{z\in\nu}[z_i]\,|1 \leq i \leq k\right), \left(\mathop{\mathbb{E}}_{z\sim\nu}[z_i z_j]\,|1 \leq i < j \leq k\right)\right)$$

of distributions ν supported on $f^{-1}(1)$. For a measure Λ on $\mathcal{C}(f)$ and a subset $S \subseteq [k]$, let Λ_S denote the projection of Λ onto the co-ordinates in S. For a permutation $\pi : S \mapsto S$ and a choice of signs $b \in \{-1,1\}^S$, let $\Lambda_{S,\pi,b}$ denote the measure Λ_S after permuting the indices in S according to π and then (possibly) negating the co-ordinates according to multiplication by $\{b_i\}_{i\in S}$. We are now ready to state our characterization.

Definition 1.1 *Let $\mathcal{A}_s$ be the family of all predicates (of all arities) $f : \{-1,1\}^k \mapsto \{0,1\}$ such*

[6] The characterization in [5] is recursively enumerable, i.e. there is a procedure that on a predicate that is approximable, terminates and declares so. Our characterization is also recursively enumerable though it is not clear from its statement and one has to work through the proof. We omit this aspect from the current version of the paper.

[7] The characterization in [5], in hindsight, may also be stated in terms of similar symmetry properties, and we do so in this paper.

that there is a probability measure Λ on $\mathcal{C}(f)$ such that for every $1 \leq t \leq k$, the signed measure

$$\Lambda^{(t)} := \mathop{\mathbb{E}}_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \left[\left(\prod_{i=1}^{t} b_i \right) \cdot \hat{f}(S) \cdot \Lambda_{S,\pi,b} \right] \tag{1.1}$$

vanishes identically. If so, Λ itself is said to vanish.

Much elaboration is in order. In the above expression, the expectation is over a random subset of $[k]$ of size t, a random permutation π of S and a random choice of signs b on S. The coefficients $\hat{f}(S)$ are the Fourier coefficients of the predicate f, namely, the coefficients in the Fourier representation:

$$f(x_1, \ldots, x_k) = \rho(f) + \sum_{S\neq\emptyset} \hat{f}(S) \prod_{i\in S} x_i.$$

A *signed measure* is allowed to take negative values as well (as is evident from the possibly negative sign of $\hat{f}(S)$ and $\prod_{i=1}^{t} b_i$ in the above expression). An equivalent way to state the condition is that if one writes the Expression (1.1) as a difference of two non-negative measures $\Lambda^{(t),1}$ and $\Lambda^{(t),2}$ by grouping the terms with positive and negative coefficients respectively, then the two non-negative measures are identical.

Our characterization states that if $f \in \mathcal{A}_s$, then f is strongly approximation resistant (UG-hardness) and otherwise weakly approximable. In the former case, the vanishing measure Λ is a *hard to round* measure (in fact any proposed hard to round measure must be a vanishing measure). In the latter case, we can in fact conclude that the predicate is weakly approximable via a natural SDP relaxation followed by a $(k+1)$-*dimensional rounding algorithm*. A $(k+1)$-dimensional rounding algorithm samples a $(k+1)$-dimensional rounding function $\psi : \mathbb{R}^{k+1} \mapsto \{-1, 1\}$ from an appropriate distribution, projects the SDP vectors onto a random $(k+1)$-dimensional subspace and then rounds using ψ. We find this conclusion rather surprising. As mentioned earlier, it follows from Raghavendra [29] that if a predicate is approximable then it is so via (the same) SDP relaxation followed by a rounding. However his rounding (and/or the one in [30]) is high dimensional in the sense that one first projects onto a random d-dimensional subspace and then rounds using an appropriately sampled function $\psi : \mathbb{R}^d \mapsto \{-1, 1\}$ and there is no a priori upper bound on the dimension d required.

It is instructive to check that our characterization generalizes the sufficient condition for approximation resistance due to Austrin and Mossel [6]. Suppose that a predicate supports a pairwise independent distribution. This amounts to saying that the $k + \binom{k}{2}$ dimensional all-zeroes vector lies in the polytope $\mathcal{C}(f)$. It is immediate that the measure Λ concentrated at this single vector is vanishing (the all-zeroes vector and its projections onto subsets S remain unchanged under sign-flips via $b \in \{-1, 1\}^S$ and these terms cancel each other out due to the sign $\prod_{i\in S} b_i$ in the expression) and hence the predicate is strongly approximation resistant. It is also instructive to check the case $t = 1$. In this case, the condition implies, in particular, that

$$\mathop{\mathbb{E}}_{\zeta\sim\Lambda} \left[\sum_{i=1}^{k} \hat{f}(\{i\}) \cdot \zeta(i) \right] = 0.$$

Here $\zeta(i)$ denotes the i^{th} first moment (i.e. bias) in the vector $\zeta \in \mathcal{C}(f)$. For all the predicates that are known to be approximation resistant so far in literature, there is always a single *hard to round point* ζ, i.e. the measure Λ is concentrated at a single point ζ. In that case, the above condition specializes to $\sum_{i=1}^{k} \hat{f}(\{i\}) \cdot \zeta(i) = 0$ and this condition is known to be necessary (as a

folklore among the experts at least). This is because otherwise a rounding that simply rounds each variable according to its bias given by the LP relaxation (and then flipping signs of all variables simultaneously if necessary) will strictly exceed the threshold $\rho(f)$. The term $\sum_{i=1}^{k} \hat{f}(\{i\}) \cdot \zeta(i)$ represents the contribution to the *advantage over* $\rho(f)$ by the level-1 Fourier coefficients and a standard trick allows one to ignore the (potentially troublesome) interference from higher order Fourier levels. The conditions for $t \geq 2$ intuitively rule out successively more sophisticated rounding strategies and taken together for all $t \in [k]$ form a complete set of necessary *and* sufficient conditions for strong approximation resistance.

It seems appropriate to point out another aspect in which our result differs from [29, 30]. It can be argued (as also discussed in [5]) that [30] also gives a characterization of approximation resistance in the following sense. The authors therein propose a brute force search over all instances and their potential SDP solutions on $N = N(\varepsilon)$ variables which determines the hardness threshold up to an additive ε. Thus if a predicate is approximable with an advantage of say 2ε over the trivial $\rho(f)$ threshold and if ε were known a priori, then the brute force search will be able to affirm this. However, there is no a priori lower bound on ε and thus this characterization is not known to be decidable either. Moreover, it seems somewhat of a stretch to call it a characterization because of the nature of the search involved. On the other hand, ours is a characterization at least in the sense that it depends purely on the predicate f and the corresponding polytope $\mathcal{C}(f)$. More specifically, the characterization does not depend on the topology (i.e. the hyper-graph structure) of the CSP instance. We find this conclusion rather surprising as well. A priori, what might make a predicate hard is both a *hard to round* measure over local LP/SDP distributions (i.e. a measure Λ on $\mathcal{C}(f)$) as well as the topology of the constraint hyper-graph (i.e. how the variables and constraints *fit together*). Our conclusion is that the latter aspect is not relevant, not in any direct manner at least. This conclusion may be contrasted against Raghavendra's result. He shows that any SDP integrality gap *instance* can be used as a gadget towards proving a UG-hardness result with the same gap. The *instance* here refers to both the variable-constraints topology and the local LP/SDP distributions and from his result, it is not clear whether one or the other or both the aspects are required to make the CSP hard.

When CSP instances are restricted to be k-partite as in [5], we are able to obtain a complete characterization. For the family $\mathcal{A}_p$ defined below, if $f \in \mathcal{A}_p$ then the partite version is strongly approximation resistant and otherwise the partite version is approximable (i.e. best of both the worlds).

Definition 1.2 *Let $\mathcal{A}_p$ be the family of all predicates (of all arities) $f : \{-1,1\}^k \mapsto \{0,1\}$ such that there is a probability measure Λ on $\mathcal{C}(f)$ such that for every $S \subseteq [k], S \neq \emptyset$, the signed measure*

$$\Lambda^S \;:=\; \hat{f}(S) \cdot \mathop{\mathbb{E}}_{b \in \{-1,1\}^S} \left[\left(\prod_{i \in S} b_i \right) \cdot \Lambda_{S,b} \right] \tag{1.2}$$

vanishes identically.

The difference from Definition 1.1 is that each non-empty set S is considered separately and there are no permutations of the set. We note that for even predicates, the first k co-ordinates in the body $\mathcal{C}(f)$ corresponding to the first moments (i.e. "biases") can be assumed to be identically zero and then the characterization boils down to one in [5] (though there it is stated differently).

We point out some directions left open by the discussion so far (we do not consider these as the focus of the current paper). Firstly, it would be great to obtain a complete characterization of

approximation resistance as opposed to that of strong resistance that we obtain and see whether the two coincide (we wouldn't venture a guess). Secondly, it would be nice to show that our characterization is decidable. Thirdly, we are not aware of an approximation resistant predicate where one needs a combination of more than one *hard to round* points in $\mathcal{C}(f)$. In other words, it might be the case that for every strongly approximation resistant predicate, there exists a vanishing measure Λ on $\mathcal{C}(f)$ that is concentrated on a single point or on a bounded number of points with an apriori bound. If this were the case, our characterization will be decidable (we omit the proof). Finally, it will be interesting to show that for some special classes of predicates our characterization takes a much simpler form. For instance, [11] asks whether there is a linear threshold predicate that is approximation resistant. It would be nice if for such predicates our characterization takes a simpler form and leads to a resolution of the question.

Results for Linear and Semidefinite Relaxations

We now move onto a discussion about our results concerning the notion of (strong) approximation resistance in the context of linear and/or semi-definite programming relaxations. A CSP instance can be formulated as an integer program and its variables may be *relaxed* to assume real values (in the case of LP relaxation) or vector values (in the case of SDP relaxation). The *integrality gap* of a relaxation is the maximum gap between the optimum of the integer program and the optimum of the relaxed program. An integrality gap instance is a concrete instance of a CSP whose LP/SDP optimum is *high* and the integer optimum is *low*. Constructing such gap instances is taken as evidence that the LP/SDP based approach will not achieve good approximation to the CSP. The LP/SDP relaxation may be *ad hoc* or may be obtained by systematically adding inequalities, in successive *rounds*, each additional round yielding a potentially *tighter* relaxation. The latter method is referred to as an LP or SDP hierarchy and several such hierarchies have been proposed and well-studied [12].

In this paper, we focus on one ad hoc relaxation that we call *basic relaxation* and two hierarchies, namely the *mixed hierarchy* and the *Sherali-Adams LP hierarchy*. We refer to Section 2 for their formal definitions, but provide a quick sketch here. Consider a CSP(f) instance with a k-ary predicate f, a set of variables $V = \{x_1, \ldots, x_n\}$ and constraints $C_1, \ldots, C_m$. We think of the number of rounds r as k or more. The r-round Sherali-Adams LP is required to provide, for every set $S \subseteq V, |S| \leq r$, a *local* distribution $D(S)$ over assignments to the set S, namely $\{-1, 1\}^S$. The local distributions must be consistent in the sense that for any two sets S, T of size at most r and $S \cap T \neq \emptyset$, the local distributions to S and T have the same marginals on $S \cap T$. The r-round mixed hierarchy, in addition, is supposed to assign unit vectors $\mathbf{u}_i$ to variables x_i such that the pairwise inner products of these vectors match the second moments of the local distributions: $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \mathbb{E}_{\sigma \sim D(\{i,j\})} [\sigma(i)\sigma(j)]$ (this is a somewhat simplified view). The *basic relaxation* is a reduced form of the k-round mixed hierarchy where a local distribution over a set S needs to be specified only if S is a set of k variables of some constraint C_ℓ. The only consistency requirements are that $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \mathbb{E}_{\sigma \sim D(S)} [\sigma(i)\sigma(j)]$ if variables i, j appear together inside some constraint C_ℓ on set S. Finally, the objective function for all three programs is the same: the probability that an assignment sampled from the local distribution over a constraint satisfies the predicate (accounting for variable negations), averaged over all constraints.

A (c, s)-integrality gap for a relaxation is an instance that is at most s-satisfiable, but has a feasible LP/SDP solution with objective value at least c. A predicate is approximation resistant w.r.t. a given relaxation if the relaxation has $(1 - o(1), \rho(f) + o(1))$ integrality gap. The general result of Raghavendra referred to before shows that for any gap location (c, s), UG-hardness is equivalent to

integrality gap for the basic relaxation. Moreover, the general results of Raghavendra and Steurer [31] and Khot and Saket [26] show that the integrality gap for basic relaxation is equivalent to that for a super-constant number of rounds of the mixed hierarchy.

All our integrality gap constructions achieve strong resistance, namely that the constructed CSP instance has LP or SDP value $1 - o(1)$ and for any (integral) assignment to the instance, the fraction of satisfied constraints is in the range $[\rho(f) - o(1), \rho(f) + o(1)]$. We call this strong $(1 - o(1), \rho(f) \pm o(1))$ integrality gap. Our characterization of strong approximation resistance for the basic relaxation and the mixed hierarchy is the same and coincides with one in Definition 1.1 whereas that for the Sherali-Adams LP is different and presented below.

When $f \in \mathcal{A}_s$ as in Definition 1.1, we construct a strong $(1 - o(1), \rho(f) \pm o(1))$ integrality gap for the basic relaxation. From the general results [29, 31, 26] mentioned before, integrality gap for basic relaxation can be translated into the same gap for mixed hierarchy and into UG-hardness. For us, in the NO case, we need to be more careful since we want to preserve the strong resistance, namely that every (integral) assignment satisfies between $\rho(f) \pm o(1)$ fraction of assignments. Still, these translations are by now standard and well-understood and are omitted from the current paper. When $f \notin \mathcal{A}_s$, we know that the predicate is weakly approximable and moreover the algorithm is a rounding of the basic relaxation. When $f \in \mathcal{A}_p$ as in Definition 1.2, the UG-hardness as well as integrality gap constructions can be ensured to be on k-partite instances, as in [5].

Finally we focus on the characterization of approximation resistance in Sherali-Adams LP hierarchy. Here the situation is fundamentally different at a conceptual level. The difference is illustrated by the (arguably the simplest) predicate 2LIN. Goemans and Williamson show that 2LIN is approximable via an SDP relaxation, namely the basic relaxation according to our terminology. In fact the approximation is really close: on an $(1-\varepsilon)$-satisfiable instance, the relaxation finds $(1 - \frac{\arccos(1-\varepsilon)}{\pi})$-satisfying assignment (which is asymptotically $1 - O(\sqrt{\varepsilon})$). It is also known that this is precisely the integrality gap as well as UG-hardness gap [16, 25]. However, the predicate turns out to be approximation resistant in the Sherali-Adams LP hierarchy as shown by de la Vega and Mathieu [13]! They show $(1 - o(1), \frac{1}{2} + o(1))$ integrality gap for $\omega(1)$ rounds of the Sherali-Adams hierarchy, which is subsequently improved to $n^{\Omega(1)}$ rounds in [10].

Even though the approximation resistance in Sherali-Adams LP hierarchy is fundamentally different, our characterization of the strong resistance here looks *syntactically* similar to the ones before, once we ignore the second moments (which are not available in the LP case).

Definition 1.3 *Let $\mathcal{A}_l$ be the family of all predicates (of all arities) $f : \{-1,1\}^k \mapsto \{0,1\}$ such that there is a probability measure Λ^* on $\mathcal{C}^*(f)$ such that for every $1 \le t \le k$, the signed measure*

$$\Lambda^{*,(t)} := \mathop{\mathbb{E}}_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \left[\left(\prod_{i=1}^{t} b_i \right) \cdot \hat{f}(S) \cdot \Lambda^*_{S,\pi,b} \right] \tag{1.3}$$

vanishes identically. Here $\mathcal{C}^*(f)$ is the projection of the polytope $\mathcal{C}(f)$ to the first k co-ordinates corresponding to the first moments and $\Lambda^*_{S,\pi,b}$ are as earlier, but for the projected polytope $\mathcal{C}^*(f)$.

We show that if $f \in \mathcal{A}_l$, then there is a strong $(1 - o(1), \rho(f) \pm o(1))$ integrality gap for a super-constant number of rounds of Sherali-Adams hierarchy. Otherwise there is a weak approximation given by k-rounds of the hierarchy. For the class of symmetric k-ary predicates, our characterization takes a simple form. If f is symmetric then $f \in \mathcal{A}_l$ if and only if there are inputs $x, y \in \{-1,1\}^k$ such that $f(x) = f(y) = 1$ and $\sum_{i=1}^k x_i \ge 0$, $\sum_{i=1}^k y_i \le 0$.

For Sherali-Adams hierarchy, we also get a complete characterization of strong approximation resistance, though not as explicit as in Definition 1.3. Also, if a predicate is weakly approximable in this hierarchy, then this is so via a generic algorithm that solves a k-round LP and then rounds every variable using its bias in the LP solution. Thus only the LP biases (and their consistency with local distributions over constraints) are *useful* towards algorithmic purpose. As far as we know, these conclusions were not known before.

1.1 Overview of the Proof Techniques

In this section we give an informal overview of the main ideas and techniques used in our results. A significant ingredient in our results is the Von Neumann min-max theorem which may have more applications in future. The theorem was also used by O'Donnell and Wu [28] towards characterizing the *approximability curve* for the MAX-CUT problem.

We first focus on the main result in the paper, namely that a predicate f is strongly approximation resistant if and only if $f \in \mathcal{A}_s$ as in Definition 1.1. We make several simplifying assumptions and use informal mathematically imprecise language as we proceed (for the sake of a cleaner overview only).

Let $f : \{-1,1\}^k \mapsto \{0,1\}$ be the predicate under consideration with $\rho(f) = \frac{|f^{-1}(1)|}{2^k}$. We make a simplifying assumption that the predicate f is even, i.e. $f(-z) = f(z)\ \forall z \in \{-1,1\}^k$. This allows us to assume that the first moments (i.e. "biases") $\mathbb{E}_{z\sim\nu}[z_i]$ are all zero for any distribution ν supported on $f^{-1}(1)$ and can be safely ignored. [8] Therefore we let the polytope $\mathcal{C}(f)$ to be the set of all $\binom{k}{2}$-dimensional *second moments vectors* $\zeta(\nu) = \left(\mathbb{E}_{z\sim\nu}[z_i z_j] \mid 1 \le i < j \le k\right)$ over all distributions ν supported on $f^{-1}(1)$. Our main concern is whether there is an efficient algorithm for CSP(f) that achieves a weak approximation, i.e. on an $1-o(1)$) satisfiable instance obtains an assignment such that the fraction of satisfied constraints is outside the range $[\rho(f) - \Omega(1), \rho(f) + \Omega(1)]$. We make the simplifying assumption that the CSP instance is in fact perfectly satisfiable. This implies that the basic relaxation yields, for every constraint C that depends on variables say $x_1, \ldots, x_k$, a distribution $\nu(C)$ over the set of satisfying assignments $f^{-1}(1)$ and unit vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ such that $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \mathbb{E}_{z\sim\nu}[z_i \cdot z_j]$. As noted, $\zeta(\nu(C))$ then is a $\binom{k}{2}$-dimensional vector of the second moments (which equal $\langle \mathbf{u}_i, \mathbf{u}_j \rangle$). The uniform distribution over the vectors $\zeta(\nu(C))$ over all constraints C is then a probability measure λ on $\mathcal{C}(f)$. We regard the measure λ as essentially representing the given CSP instance (a priori, we seem to be losing information by ignoring the topology of the instance, but as we will see this doesn't matter).

Note that in the relaxed solution, the vector assignment is global in the sense that the vector assigned to each CSP variable is fixed, independent of the constraint C in which the variable participates in whereas the distribution $\nu(C)$ is local in the sense that it depends on the specific constraint C.

Our main idea is to propose a family of algorithms based on "d-dimensional roundings" of the SDP solution for $d = k+1$ and to show that either one such algorithm achieves a weak approximation or else the polytope $\mathcal{C}(f)$ supports a probability measure Λ as in Definition 1.1 (note again that we are ignoring the first moments). In the latter case, the existence and symmetry of Λ leads naturally to a strong $(1-o(1), \rho(f) \pm o(1))$ integrality gap for the basic relaxation (and therefore mixed hierarchy) and a UGC-hardness result for GapCSP$(f)_{1-o(1),\rho(f)\pm o(1)}$, showing that the predicate is strongly approximation resistant.

[8] See [5] for elaboration where the same assumption is used.

The proposed family of d-dimensional roundings is easy to describe: any function $\psi : \mathbb{R}^d \mapsto \{-1, 1\}$ serves as a candidate rounding algorithm where the SDP vectors $\{\mathbf{u}_i\}$ are projected onto a random d-dimensional subspace inducing $\mathbf{u}_i \mapsto \mathbf{y}_i \in \mathbb{R}^d$ and then the i^{th} variable is assigned a boolean value $\psi(\mathbf{y}_i)$. From the algorithmic viewpoint, one seeks a rounding function ψ (more generally a distribution over ψ) such that its "performance" on *every* instance λ[9] significantly deviates from $\rho(f)$ (in average, if a distribution over ψ is used). From the hardness viewpoint, a natural goal then would be to come up with a "hard-to-round measure" λ on $\mathcal{C}(f)$ such that the "performance" of *every* rounding function ψ is within $\rho(f) \pm o(1)$.

These considerations lead naturally to a two-player zero-sum game between Harry, the "hardness player" and Alice, the "algorithm player" (we view Harry as the row player and Alice as the column player). The pure strategies of Harry are the probability measures λ on $\mathcal{C}(f)$ to be rounded and the pure strategies of Alice are the rounding functions $\psi : \mathbb{R}^d \mapsto \{-1, 1\}$. The payoff to Alice when the two players play (λ, ψ) respectively is the "deviation from $\rho(f)$" achieved by rounding λ using ψ. More precisely, consider the scenario where the set of local distributions on CSP constraints is represented by the measure λ. The local distribution on a randomly selected constraint is a sample $\zeta \sim \lambda$ along with vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ whose pairwise inner products match ζ. During the rounding process, the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are projected onto a random d-dimensional subspace, generating a sequence of k points $\mathbf{y}_1, \ldots, \mathbf{y}_k \in \mathbb{R}^d$ that are standard d-dimensional Gaussians with correlations ζ. The CSP variables are then rounded to boolean values $\psi(\mathbf{y}_1), \ldots, \psi(\mathbf{y}_k)$. Whether these values satisfy the constraint or not is determined by plugging them in the Fourier representation of the predicate f. The "deviation from $\rho(f)$" is precisely this Fourier expression without the constant term (which is $\rho(f)$). Given this intuition, we define the payoff to Alice as the expression:

$$\mathsf{PayOff}(\lambda, \psi) \;:=\; \left| \mathop{\mathbb{E}}_{\zeta \sim \lambda} \; \mathop{\mathbb{E}}_{\mathbf{y}_1, \ldots, \mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\sum_{S \neq \emptyset} \hat{f}(S) \cdot \prod_{i \in S} \psi(\mathbf{y}_i) \right] \right|, \tag{1.4}$$

where $\mathcal{N}_d(\zeta)$ denotes a sequence of k standard d-dimensional Gaussians with correlations ζ. We apply Von Neumann's min-max theorem and conclude that there exists a number L, namely the "value" of the game, a mixed equilibrium strategy Γ (a distribution over ψ) for Alice and an equilibrium strategy Λ (a pure one as we will observe!) for Harry. Actually Von Neumann's theorem applies only to games where the sets of strategies for both players are finite, but we ignore this issue for now. Depending on whether the value of the game L is strictly positive or zero (it is non-negative since the payoff is non-negative), we get the "dichotomy" that the predicate f is weakly approximable or strongly approximation resistant (modulo UGC).

The conclusion when $L > 0$ is easy: in this case Alice has a mixed strategy Γ such that her payoff (expected over Γ) is at least L for *every* pure strategy λ of Harry. This is same as saying that if a rounding function $\psi \sim \Gamma$ is sampled and then used to round the relaxed solution, it achieves a deviation L from $\rho(f)$ for every CSP instance λ.

The conclusion when $L = 0$ is more subtle: in this case in general Harry has a mixed strategy, say $\mathcal{D}$, such that for *every* pure strategy ψ of Alice, her expected payoff (expected over $\lambda \sim \mathcal{D}$) is zero. We observe that Harry may replace his mixed strategy $\mathcal{D}$ by a pure strategy Λ. Noting that $\mathcal{D}$ is a distribution over measures λ, we let Λ be the single *averaged* measure informally written as $\Lambda := \mathbb{E}_{\lambda \sim \mathcal{D}}[\lambda]$. Since for every ψ, Expression (1.4) is supposed to be zero averaged over $\lambda \sim \mathcal{D}$, it must also be zero for Λ itself! Thus we conclude that for the measure Λ over $\mathcal{C}(f)$, for every

[9] We recall again that for any CSP instance, λ is the uniform distribution over $\zeta(\nu(C))$ over all constraints C and thus a probability measure on $\mathcal{C}(f)$. The measure λ now represents the whole instance.

$\psi : \mathbb{R}^d \mapsto \{-1, 1\}$:

$$\mathop{\mathbb{E}}_{\zeta\sim\Lambda} \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} \left[\sum_{S\neq\emptyset} \hat{f}(S)\cdot\prod_{i\in S}\psi(\mathbf{y}_i)\right] \;=\; 0. \tag{1.5}$$

Now we view this expression as a polynomial in (uncountable number of) variables $\{\psi(\mathbf{y}) \mid y \in \mathbb{R}^d\}$. Since the polynomial is identically zero, we may equate every coefficient of this polynomial to zero.

Fix any $1 \le t \le k$. For every $\mathbf{y}_1, \ldots, \mathbf{y}_t \in \mathbb{R}^d$, we are interested in the coefficient of the monomial $\prod_{i=1}^t \psi(\mathbf{y}_i)$. Firstly, this coefficient can arise from precisely the sets S with $|S| = t$. Secondly, for a fixed set $S, |S| = t$, the coefficient is really the joint density of t standard d-dimensional Gaussians with correlations ζ_S at the sequence $(\mathbf{y}_1, \ldots, \mathbf{y}_t)$, where ζ_S is same as ζ restricted to indices in S. Thirdly, for any permutation $\pi : [t] \mapsto [t]$, we must consider all sequences $(y_{\pi(1)}, \ldots, y_{\pi(t)})$ and add up their coefficients (i.e. Gaussian densities) since they all correspond to the same monomial $\prod_{i=1}^t \psi(\mathbf{y}_i)$. Finally, we did not mention this so far, but we need to allow only odd rounding functions ψ, i.e. $\psi(-\mathbf{y}) = -\psi(\mathbf{y})$, to account for the issue of variable negations in CSPs. This has the effect that the monomials $\prod_{i=1}^t \psi(b_i \cdot \mathbf{y}_i)$ are same as $\prod_{i=1}^t b_i \cdot \prod_{i=1}^t \psi(\mathbf{y}_i)$ for a choice of signs $b_i \in \{-1, 1\}$, and hence their coefficients (i.e. Gaussian densities) must be added up together. With all these considerations, the coefficient of the monomial $\prod_{i=1}^t \psi(y_i)$ can be written as:

$$\mathop{\mathbb{E}}_{\zeta\sim\Lambda}\left[\sum_{S,|S|=t}\;\sum_{\pi:[t]\mapsto[t]}\;\sum_{b\in\{-1,1\}^t} \hat{f}(S)\cdot\left(\prod_{i=1}^t b_i\right)\gamma_{t,d}\big((\mathbf{y}_1,\ldots,\mathbf{y}_t),\zeta_{S,\pi,b}\big)\right].$$

Here $\zeta_{S,\pi,b}$ is the sequence of correlations between the indices in S after accounting for the permutation of indices according to π and the sign-flips according to $b \in \{-1, 1\}^t$. Also $\gamma_{t,d}\big((\mathbf{y}_1, \ldots, \mathbf{y}_t), \xi\big)$ is the joint density of t standard d-dimensional Gaussians with correlations ξ. Defining the "signed measure" $\Lambda^{(t)}$ as in Equation (1.1), the conclusion that the above coefficient is zero (for every $(\mathbf{y}_1, \ldots, \mathbf{y}_t)$), can be written as:

$$\forall \mathbf{y}_1, \ldots, \mathbf{y}_t \in \mathbb{R}^d, \quad \int \gamma_{t,d}\big((\mathbf{y}_1, \ldots, \mathbf{y}_t), \xi\big)\; d\Lambda^{(t)}(\xi) = 0.$$

In words, w.r.t. the signed measure $\Lambda^{(t)}$ on $[-1, 1]^{\binom{t}{2}}$ (corresponding to all possible correlation vectors between t standard 1-dimensional Gaussians), the integral of every function $\gamma_{t,d}((\mathbf{y}_1, \ldots, \mathbf{y}_t), \cdot)$ vanishes (there is one such function for every fixed choice of $(\mathbf{y}_1, \ldots, \mathbf{y}_t)$). The class of these functions is rich enough that we are able to conclude that the signed measure $\Lambda^{(t)}$ itself must identically vanish.

This proves the existence of the measure Λ as in Definition 1.1. After this, the construction of the strong $(1 - o(1), \rho(f) + \pm o(1))$ integrality gap for the CSP is obtained by generalizing the construction for MAX-CUT due to Feige and Schechtman [16]. We describe the construction in the continuous setting and ignore the discretization step. The variables in the CSP instance correspond to points in $\mathbb{R}^N$ for a high enough dimension N and the variables for $\mathbf{y}$ and $-\mathbf{y}$ are designated as negations of each other. The constraints of the CSP are defined by sampling $\zeta \sim \Lambda$ and then sampling k Gaussian points $\mathbf{y}_1, \ldots, \mathbf{y}_k \in \mathbb{R}^N$ with correlations ζ and placing a constraint on these variables. For the completeness part, one observes that for large N the space $\mathbb{R}^N$ with the Gaussian measure is (up to $o(1)$ errors) same as the unit sphere $\mathbb{S}^{N-1}$ towards our purpose and we may assume that all the CSP variables lie on the unit sphere. Each point on the sphere is assigned a vector that is itself and for every constraint, the local distribution equals ν if $\zeta = \zeta(\nu)$ is used towards that constraint. For the soundness part, an assignment to the CSP corresponds to a function

$\psi : \mathbb{R}^N \mapsto \{-1, 1\}$ and the "deviation from $\rho(f)$" is precisely the expression in Equation (1.4), if $y_1, \ldots, y_k$ were chosen from $\mathbb{R}^N$ instead of $\mathbb{R}^d$ ($d = k+1$ therein). The symmetry property of Λ (i.e. that the signed measure $\Lambda^{(t)}$ vanishes for every $1 \leq t \leq k$) ensures that this expression vanishes identically and hence no CSP assignment can deviate from $\rho(f)$. We would like to emphasize here that the existence of Λ was deduced only assuming that no $(k+1)$-dimensional rounding deviates from $\rho(f)$, but once existence of Λ is established, it automatically implies that no higher dimensional rounding deviates from $\rho(f)$ either!

Once the integrality gap is established, the UGC-hardness of $\mathrm{GapCSP}(f)_{(1-o(1),\rho(f)\pm o(1))}$ follows almost automatically from the general result of Raghavendra and the same integrality gap for a super-constant number of rounds of the mixed hierarchy follows almost automatically from the general results of Raghavendra and Steurer [31], and Khot and Saket [26] (some care is required).

As we said, this is a simplified and informal view and we actually need to work around all the simplifying assumptions we made, formalize all the arguments, and address many issues that we hid under the carpet, e.g. setting $d = k + 1$ and the reason say $d = 1$ does not work, handling the first moments, handling the possibility that a Gaussian density is degenerate, etc. Also, we cannot apply Von Neumann's min-max theorem to infinite games. In principle, one might be able to use min-max theorems for infinite games such as Glicksberg's theorem, but then one has to ensure that the strategy spaces are compact. Instead, we find it easier to work with a sequence of finite approximations to the infinite game and then use limiting arguments everywhere (this is easier said than done and this is where much of the work lies in).

We briefly state why we get a characterization of only strong approximation resistance and not approximation resistance itself. If we were to apply the same technique towards approximation resistance, we would define the payoff function in Equation (1.4) without the absolute value and then in Equation (1.5) we would conclude that the expression is upper bounded by zero (instead of concluding that its absolute value is zero and hence it is equal to zero). But then we would be stuck since if a polynomial in $[-1, 1]$-valued variables stays upper bounded by zero, we cannot necessarily conclude that it is identically zero, e.g. consider the polynomial $-x^2$. The conclusion however holds if the polynomial is multi-linear (which we leave as an exercise) or has only odd degree monomials (since flipping the sign of all variables flips the sign of the polynomial)[10]. We note that the polynomial in Equation (1.5) *appears* to be multi-linear in the following sense. The monomial $\prod_{i=1}^{t} \psi(\mathbf{y}_i)$ fails to be multi-linear only when $\mathbf{y}_i = \mathbf{y}_j$ for some $i \neq j$ which is a measure zero event. However when one formalizes the argument via discretization, say by dividing $\mathbb{R}^d$ into tiny cells, one always has a non-zero probability that $\mathbf{y}_i, \mathbf{y}_j, i \neq j$ fall into the same cell. At first sight, one might consider this issue minor and avoidable, but quite possibly it is indeed a serious issue and in the end, the characterization for approximation resistance is different from that of strong resistance. We leave this as an exciting open question.

Strong Approximation Resistance for LP Hierarchies

Now we give an overview of the characterization of strong approximation resistance (i.e. Definition 1.3) for a super-constant number of rounds of Sherali-Adams LP. We proceed along a similar line as earlier with one difference: we work with a different body $\tilde{\mathcal{C}}(f)$ instead of $\mathcal{C}(f)$.

In the LP case, the second moments are not available at all and the first moments are all one has. We will nevertheless pretend that the second moments are available by using their *dummy* setting.

[10] This is consistent with the earlier observation that for odd predicates, the notions of approximation resistance and strong approximation resistance are equivalent.

For any distribution ν supported on $f^{-1}(1)$, let the vector $\zeta = \zeta(\nu)$ consist of the k first moments $\zeta(i) = \mathbb{E}_{z\sim\nu}[z_i]$ and in addition, *dummy* second moments corresponding to those of k *independent* unit ℓ_2-norm Gaussians $g_1, \ldots, g_k$ with the given first moments, i.e. $\mathbb{E}[g_i] = \zeta(i)$ and $\mathbb{E}[g_i^2] = 1$. The body $\tilde{\mathcal{C}}(f)$ is defined as the set of all vectors $\zeta(\nu)$ over all distributions ν supported on $f^{-1}(1)$. Note that $\tilde{\mathcal{C}}(f)$ is different than the polytope $\mathcal{C}(f)$ and not necessarily convex (we never used convexity), but its projection onto the first k co-ordinates is the same as that of $\mathcal{C}(f)$, namely $\mathcal{C}^*(f)$ as in Definition 1.3.

Once the polytope $\mathcal{C}(f)$ is replaced by the body $\tilde{\mathcal{C}}(f)$, our argument proceeds as before. Note that since the second moments reflect independent Gaussians, our rounding is really using only the first moments, as ought to be the case with LPs. We conclude that either the predicate is weakly approximable or there is a probability measure Λ on $\tilde{\mathcal{C}}(f)$ that satisfies characterization in Definition 1.1. Projecting Λ onto the first k co-ordinates gives a measure Λ^* on $\mathcal{C}^*(f)$ satisfying the characterization in Definition 1.3.

Once the existence of Λ^* is established, we proceed to constructing the strong $(1-o(1), \rho(f)\pm o(1))$ integrality gap in the Sherali-Adams hierarchy. This step however turns out to be more involved than before since general results as in [29, 31, 26] are not available in the LP setting. Instead, we are able to rework the MAX-CUT construction of de la Vega and Kenyon [13] for any predicate $f \in \mathcal{A}_l$.

An intuitive way of looking at the construction is as follows. The variables of the CSP are points in the interval $[-1, 1]$ and the variables for x and $-x$ are negations of each other (called *folding*). Constraints are defined by sampling $\zeta \sim \Lambda^*$ and then placing the constraint on variables $(\zeta(1), \ldots, \zeta(k))$. The local distribution for this constraint is ν such that $\zeta = \zeta(\nu)$. The LP-bias of a variable x is x itself. The vanishing condition in Definition 1.3 implies that any (measurable) $\{-1, 1\}$-assignment to this CSP instance satisfies exactly $\rho(f)$ fraction (measure) of the constraints. This conclusion also holds for $[-1, 1]$-valued assignments appropriately interpreted.

This continuous instance only has a *basic* LP solution, i.e. the local distributions are defined only for constraints. We now construct the actual instance as follows. We discretize the interval $[-1, 1]$ by picking equally spaced points $x_1, \ldots, x_s$ with fine enough granularity (and ensuring that a point and its negation are both included and are *folded*). Each variable x_i is now *blown up* into a block of n/s variables for a large n (so the total number of variables is n). Whenever a constraint is generated in the continuous setting by sampling $\zeta \sim \Lambda^*$, we first round $\zeta(j)$ to nearest x_{i_j} and then the constraint is actually placed on randomly chosen variables from blocks corresponding to $x_{i_1}, \ldots, x_{i_k}$ respectively. This is the way *one* constraint is randomly introduced and the process is repeated independently m times for $m \gg n$. This defines the CSP instance as a k-uniform hyper-graph. By deleting a small fraction of the constraints, one ensures that the hyper-graph has super-constant girth. Finally, de la Vega and Kenyon [13] construction is reworked to construct local distributions for all r-sets of variables, i.e. for the r-round Sherali-Adams LP. Our presentation is somewhat different than that in [13]: we find it easier to first construct a nearly correct LP solution and then correct it as in [31, 26].

One interesting and novel feature of our construction is how the CSP instance is constructed and how the "soundness" is proved as opposed to a standard construction of random CSPs.

A standard construction, in one step, generates a constraint by uniformly selecting a k-subset of variables and then randomly selecting the *polarities* (i.e. whether a variable occurs in a negated form or not). This step is then repeated independently to generate $m \gg n$ constraints. Since the polarities are randomly chosen in each step, for any fixed global assignment, the probability that the assignment satisfies the constraint is precisely $\rho(f)$, and then one uses the Chernoff bound and

the union bound to conclude that w.h.p. every global assignment to the instance satisfies between $\rho(f) \pm o(1)$ fraction of the constraints.

In our case, the one step of generating a constraint is different. In particular, the k-subset of variables chosen is not necessarily uniformly random (it depends on Λ^* since $\zeta \sim \Lambda^*$) and the polarities are not necessarily random either (they depend on signs of $\zeta(1), \ldots, \zeta(k)$ due to *folding*). However it is still true that for any fixed global assignment, the probability that the assignment satisfies the constraint is precisely $\rho(f)$ (up to $o(1)$ errors introduced by discretization)! This property is simply inherited from the continuous setting by viewing the global assignment as a function $\psi : \{x_1, \ldots, x_s\} \mapsto [-1, 1]$ where $\psi(x_i)$ is the average of the global values to variables in block x_i! This concludes our overview.

2 Preliminaries and Our Results

In this section, we present formal definitions and statements of our results and a preliminary background on mathematical tools used.

2.1 Constraint Satisfaction Problems

Definition 2.1 *For a predicate $f : \{-1, 1\}^k \to \{0, 1\}$, an instance Φ of $\mathsf{CSP}(f)$ consists of a set of variables $\{x_1, \ldots, x_n\}$ and a set of constraints $C_1, \ldots, C_m$ where each constraint C_i is over a k-tuple of variables $\{x_{i_1}, \ldots, x_{i_k}\}$ and is of the form*

$$C_i \equiv f(x_{i_1} \cdot b_{i_1}, \ldots, x_{i_k} \cdot b_{i_k})$$

for some $b_{i_1}, \ldots, b_{i_k} \in \{-1, 1\}$. For an assignment $A : \{x_1, \ldots, x_n\} \mapsto \{-1, 1\}$, let $\mathsf{sat}(A)$ denote the fraction of constraints satisfied by A. The instance is called α-satisfiable if there exists an assignment A such that $\mathsf{sat}(A) \geq \alpha$. The maximum fraction of constraints that can be simultaneously satisfied is denoted by $\mathsf{OPT}(\Phi)$, i.e.

$$\mathsf{OPT}(\Phi) = \max_{A:\{x_1,\ldots,x_n\}\mapsto\{-1,1\}} \mathsf{sat}(A).$$

The density of the predicate is $\rho(f) = \frac{|f^{-1}(1)|}{2^k}$.

For a constraint C of the above form, we use x_C to denote the tuple of variables $(x_{i_1}, \ldots, x_{i_k})$ and b_C to denote the tuple of bits $(b_{i_1}, \ldots, b_{i_k})$. We then write the constraint as $f(x_C \cdot b_C)$. We also denote by S_C the set of indices $\{i_1, \ldots, i_k\}$ of the variables participating in the constraint C.

Definition 2.2 *A predicate $f : \{-1, 1\}^k \to \{0, 1\}$ is called* approximable *if there exists a constant $\varepsilon > 0$ and a polynomial time algorithm, possibly randomized, that given an $(1-\varepsilon)$-satisfiable instance of $\mathsf{CSP}(f)$, outputs an assignment A such that $\mathbb{E}_A[\ \mathsf{sat}(A)\ ] \geq \rho(f) + \varepsilon$. Here the expectation is over the randomness used by the algorithm. The predicate is called* weakly approximable *if the output of the algorithm deviates from $\rho(f)$ in expectation, i.e. $\mathbb{E}_A[\ |\mathsf{sat}(A) - \rho(f)|\ ] \geq \varepsilon$.*

We define the notions of approximation resistance and strong approximation resistance. Towards this, it is convenient to define gap versions of the CSP. Though the gap versions can be defined w.r.t. any *gap location*, we do so only for the location that is of interest to us, namely $1 - o(1)$

versus $\rho(f) \pm o(1)$. We say that a decision problem is UG-hard if there is polynomial time reduction from the Unique Games Problem [24] to the problem under consideration (we will not be directly concerned with the Unique Games Problem and the Conjecture; hence their discussion is deferred to the end of the preliminaries section).

Definition 2.3 *Let $\varepsilon > 0$ be a constant.*

Let $\mathsf{GapCSP}(f)_{1-\varepsilon,\ \rho(f)+\varepsilon}$ denote the promise version of $\mathsf{CSP}(f)$ where the given instance Φ is promised to have either $\mathsf{OPT}(\Phi) \geq 1-\varepsilon$ or $\mathsf{OPT}(\Phi) \leq \rho(f)+\varepsilon$. The predicate is called approximation resistant *if for every $\varepsilon > 0$, $\mathsf{GapCSP}(f)_{1-\varepsilon,\ \rho(f)+\varepsilon}$ is UG-hard.*

Let $\mathsf{GapCSP}(f)_{1-\varepsilon,\ \rho(f)\pm\varepsilon}$ denote the promise version of $\mathsf{CSP}(f)$ where the given instance Φ is promised to have either $\mathsf{OPT}(\Phi) \geq 1-\varepsilon$ or that for every *assignment A, $|\mathsf{sat}(A) - \rho(f)| \leq \varepsilon$. The predicate is called* strongly approximation resistant *if for every $\varepsilon > 0$, $\mathsf{GapCSP}(f)_{1-\varepsilon,\ \rho(f)\pm\varepsilon}$ is UG-hard.*

2.2 The LP and SDP Relaxations for Constraint Satisfaction Problems

Below we present three LP and SDP relaxations for the $\mathsf{CSP}(f)$ problem that are relevant in this paper: the Sherali-Adams LP relaxation, mixed LP/SDP relaxation and finally the *basic relaxation.*

We start with the r-round Sherali-Adams relaxation. The intuition behind it is the following. Note that an integer solution to the problem can be given by an assignment $A : [n] \to \{-1,1\}$. Using this, we can define $\{0,1\}$-valued variables $x_{(S,\alpha)}$ for each $S \subseteq [n], 1 \leq |S| \leq r$ and $\alpha \in \{-1,1\}^S$, with the intended solution $x_{(S,\alpha)} = 1$ if $A(S) = \alpha$ and 0 otherwise. We also introduce a variable $x_{(\emptyset,\emptyset)}$, which equals 1. We relax the integer program and allow variables to take real values in $[0,1]$. Now the variables $\{x_{(S,\alpha)}\}_{\alpha\in\{-1,1\}^k}$ give a probability distribution over assignments to S. We can enforce consistency between these *local* distributions by requiring that for $T \subseteq S$, the distribution over assignments to S, when marginalized to T, is precisely the distribution over assignments to T. The relaxation is shown in Figure 1.

$$\begin{aligned}
\text{maximize} \quad & \mathop{\mathbb{E}}_{C\in\Phi}\left[\sum_{\alpha\in\{-1,1\}^k} f(\alpha\cdot b_C)\cdot x_{(S_C,\alpha)}\right] \\
\text{subject to} \quad & \\
\sum_{\substack{\alpha\in\{-1,1\}^S \\ \alpha|_T=\beta}} x_{(S,\alpha)} \;&=\; x_{(T,\beta)} && \forall T\subseteq S\subseteq[n], |S|\leq r,\ \forall\beta\in\{-1,1\}^T \\
x_{(S,\alpha)} \;&\geq\; 0 && \forall S\subseteq[n], |S|\leq r,\ \forall\alpha\in\{-1,1\}^S \\
x_{(\emptyset,\emptyset)} \;&=\; 1 &&
\end{aligned}$$

Figure 1: r-round Sherali-Adams LP for $\mathsf{CSP}(f)$

We can further strengthen the integer program by adding the quadratic constraints

$$x_{(\{i_1,i_2\},(b_1,b_2))} \;=\; x_{(\{i_1\},b_1)} \cdot x_{(\{i_2\},b_2)}\,.$$

As solving quadratic programs is NP-hard we then relax these quadratic constraints to the existence of vectors $\mathbf{v}_{(i,b)}$ and a unit vector $\mathbf{v}_{(\emptyset,\emptyset)}$, and impose the above constraints on inner products of the

corresponding vectors. Adding these SDP variables and constraints to the r-round Sherali-Adams LP as above yields the r-round mixed relaxation as in Figure 2.

$$
\begin{aligned}
\text{maximize} \quad & \mathop{\mathbb{E}}_{C\in\Phi}\left[\sum_{\alpha\in\{-1,1\}^k} f(\alpha\cdot b_C)\cdot x_{(S_C,\alpha)}\right] & & \\
\text{subject to} & & & \\
\left\langle \mathbf{v}_{(i,1)}, \mathbf{v}_{(i,-1)}\right\rangle &= 0 & & \forall i\in[n] \\
\mathbf{v}_{(i,1)} + \mathbf{v}_{(i,-1)} &= \mathbf{v}_{(\emptyset,\emptyset)} & & \forall i\in[n] \\
x_{(\{i_1,i_2\},(b_1,b_2))} &= \left\langle \mathbf{v}_{(i_1,b_1)}, \mathbf{v}_{(i_2,b_2)}\right\rangle & & \forall i_1\neq i_2\in[n], b_1,b_2\in\{-1,1\} \\
\sum_{\substack{\alpha\in\{-1,1\}^S\\ \alpha|_T=\beta}} x_{(S,\alpha)} &= x_{(T,\beta)} & & \forall T\subseteq S\subseteq[n], |S|\le r,\ \forall\beta\in\{-1,1\}^T \\
x_{(S,\alpha)} &\ge 0 & & \forall S\subseteq[n], |S|\le r,\ \forall\alpha\in\{-1,1\}^S \\
\left\|\mathbf{v}_{(\emptyset,\emptyset)}\right\|^2 &= x_{(\emptyset,\emptyset)} = 1 & &
\end{aligned}
$$

Figure 2: r-round Mixed Relaxation for $\mathsf{CSP}(f)$

Finally, the basic relaxation is a reduced form of the above mixed relaxation where only those variables $x_{(S,\alpha)}$ are included for which $S = S_C$ is the set of CSP variables for some constraint C. The consistency constraints between pairs of vectors are included only for those pairs that occur inside some constraint. The relaxation (after a minor rewriting) is shown in Figure 3.

$$
\begin{aligned}
\text{maximize} \quad & \mathop{\mathbb{E}}_{C\in\Phi}\left[\sum_{\alpha\in\{-1,1\}^k} f(\alpha\cdot b_C)\cdot x_{(S_C,\alpha)}\right] & & \\
\text{subject to} & & & \\
\left\langle \mathbf{v}_{(i,1)}, \mathbf{v}_{(i,-1)}\right\rangle &= 0 & & \forall i\in[n] \\
\mathbf{v}_{(i,1)} + \mathbf{v}_{(i,-1)} &= \mathbf{v}_{(\emptyset,\emptyset)} & & \forall i\in[n] \\
\left\|\mathbf{v}_{(\emptyset,\emptyset)}\right\|^2 &= 1 & & \\
\sum_{\substack{\alpha\in\{-1,1\}^{S_C}\\ \alpha(i_1)=b_1,\alpha(i_2)=b_2}} x_{(S_C,\alpha)} &= \left\langle \mathbf{v}_{(i_1,b_1)}, \mathbf{v}_{(i_2,b_2)}\right\rangle & & \forall C\in\Phi, i_1\neq i_2\in S_C, b_1,b_2\in\{-1,1\} \\
x_{(S_C,\alpha)} &\ge 0 & & \forall C\in\Phi,\ \forall\alpha\in\{-1,1\}^{S_C}
\end{aligned}
$$

Figure 3: Basic Relaxation for $\mathsf{CSP}(f)$

For an LP/SDP relaxation of CSP, and for a given instance Φ of the problem, we denote by $\mathsf{FRAC}(\Phi)$ the LP/SDP (fractional) optimum. For the particular instance Φ, the integrality gap is defined as $\mathsf{FRAC}(\Phi)/\mathsf{OPT}(\Phi)$. The integrality gap of the relaxation is the supremum of integrality gaps over all instances. The integrality gap thus defined is in terms of a ratio whereas we are concerned with the specific *gap location* $1-o(1)$ versus $\rho(f)+o(1)$ and also with the *strong integrality gap* as defined below.

Definition 2.4 *Let $\varepsilon > 0$ be a constant.*

A relaxation is said to have a $(1-\varepsilon, \rho(f)+\varepsilon)$*-integrality gap if there exists a CSP instance* Φ *such that* $\mathsf{FRAC}(\Phi) \geq 1-\varepsilon$ *and* $\mathsf{OPT}(\Phi) \leq \rho(f)+\varepsilon$*.*

The relaxation is said to have a strong $(1-\varepsilon, \rho(f)\pm\varepsilon)$*-integrality gap if there exists a CSP instance* Φ *such that* $\mathsf{FRAC}(\Phi) \geq 1-\varepsilon$ *and for* every *assignment* A *to the instance,* $|\mathsf{sat}(A) - \rho(f)| \leq \varepsilon$*.*

We will use known results showing that the integrality gap for the basic relaxation as in Figure 3 implies a UG-hardness result as well as integrality gap for the mixed relaxation as in Figure 2 for a super-constant number of rounds, while essentially preserving the gap. The first implication is by Raghavendra [29] and the second by Raghavendra and Steurer [31] and Khot and Saket [26]. We state these results in a form suitable for our purpose (in particular making an additional observation that the results also apply in our setting of the *strong* integrality gap).

Theorem 2.5 [29] *Let* $\varepsilon > 0$ *be an arbitrarily small constant.*

If the basic relaxation as in Figure 3 has a $(1-\varepsilon, \rho(f)+\varepsilon)$*-integrality gap, then* $\mathsf{GapCSP}(f)_{1-2\varepsilon,\rho(f)+2\varepsilon}$ *is UG-hard. Moreover, if the gap is a strong* $(1-\varepsilon, \rho(f)\pm\varepsilon)$*-gap, then* $\mathsf{GapCSP}(f)_{1-2\varepsilon,\rho(f)\pm2\varepsilon}$ *is UG-hard.*

Theorem 2.6 [31, 26] *Let* $\varepsilon > 0$ *be an arbitrarily small constant.*

If the basic relaxation as in Figure 3 has a $(1-\varepsilon, \rho(f)+\varepsilon)$*-integrality gap, then the mixed relaxation as in Figure 2 has a* $(1-2\varepsilon, \rho(f)+2\varepsilon)$*-integrality gap for a super-constant number of rounds. The same holds for a strong* $(1-\varepsilon, \rho(f)\pm\varepsilon)$*-integrality gap.*

2.3 Measure Theory and Probability

We provide a basic background on relevant tools from measure theory and probability. For further reference, please see [37, 27].

Measures, Weak* Convergence and Signed Measures

Definition 2.7 *Given a set* X *along with a* σ*-algebra* $\mathcal{F}$ *(i.e. a non-empty collection of subsets of* X *that is closed under complementation and countable union), a* measure *on* X *is a function* $m : \mathcal{F} \to [0, \infty]$ *satisfying:*

- $m(\emptyset) = 0$.
- *If* $\{R_j\}_{j=1}^{\infty} \subseteq \mathcal{F}$ *is a countable collection of disjoint sets, then* $m\left(\cup_{j=1}^{\infty} R_j\right) = \sum_{j=1}^{\infty} m(R_j)$.

We will consider only finite measures, i.e. those with $m(X) < \infty$. In particular, we will be interested in *probability measures*, i.e. those with $m(X) = 1$. The class of all probability measures on X is denoted as $\mathsf{Prob}(X)$.

We note that the Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}^n$ is the smallest σ-algebra that contains all open balls w.r.t. the standard Euclidean metric (and the sets in $\mathcal{B}$ are called measurable). It can be restricted to $X \subseteq \mathbb{R}^n$ leading to the induced σ-algebra on X, which will be the σ-algebra under consideration below. We note also that for subsets of $\mathbb{R}^n$, being compact is same as being closed and bounded via the Heine-Borel Theorem. It is also equivalent to being sequentially compact (existence of a convergent subsequene for every infinite sequence) by the Bolzano-Weierstrass Theorem.

We state the main measure-theoretic result that we need in a form convenient to us:

Theorem 2.8 *Let $X \subseteq \mathbb{R}^n$ be a compact set and $\{\Lambda_i\}_{i=1}^{\infty}$ be a sequence of probability measures on X. Then there exists a sub-sequence $\{\Lambda_{i_j}\}_{j=1}^{\infty}$ and a probability measure Λ on X such that for any continuous function $h : X \mapsto \mathbb{R}$,*

$$\lim_{j\to\infty} \int h \, d\Lambda_{i_j} = \int h \, d\Lambda. \tag{2.1}$$

This statement follows from the theorem stated below:

Theorem 2.9 (Corollary 13.9 in [37]) *Let X be a compact metric space. Then the class of probability measures $\mathsf{Prob}(X)$ is compact and metrizable in the weak* topology.*

In words, the class $\mathsf{Prob}(X)$ can be endowed with a suitable metric so that the metric topology coincides with the weak* topology. Since $\mathsf{Prob}(X)$ is compact and metrizable, it is also sequentially compact, i.e. every sequence has a convergent subsequence. The convergence is w.r.t. the metric defined on $\mathsf{Prob}(X)$ and as mentioned, this is same as the convergence in the so-called weak* topology. The latter, by definition, is precisely the statement that Equation (2.1) holds for every continuous function $h : X \mapsto \mathbb{R}$.

Let $X \subseteq \mathbb{R}^n, X' \subseteq R^{n'}$ be compact, Λ be a measure on X and $\varphi : X \mapsto X'$ be continuous. The measure $\varphi(\Lambda)$ on X' is defined in a natural way as $\varphi(\Lambda)(A') = \Lambda(\varphi^{-1}(A'))$. We will use this observation in two settings: (1) when φ is a projection of X onto a subset of co-ordinates $S \subseteq [n]$, the measure on $\mathbb{R}^{|S|}$ so obtained will be denoted as Λ_S and refereed to as the projected measure. (2) when φ is a bijection, we can pass back and forth between Λ and $\varphi(\Lambda)$, regarding them as essentially the same.

Sometimes we will describe the construction of the measure $\varphi(\Lambda)$ as above by informally saying "sample $x \sim \Lambda$ and take (or apply) $\varphi(x)$". When $h : X \mapsto \mathbb{R}$ is a real valued function, we will informally write $\mathbb{E}_{x\in\Lambda}[h(x)]$, the "expectation of $h(x)$ when x is sampled from Λ", to denote $\int h \, d\Lambda$.

We will also need the notion of a signed measure, which is a generalization of the usual (non-negative) measure.

Definition 2.10 *Given a set X along with a σ-algebra $\mathcal{F}$, a* signed measure *on X is a function $m : \mathcal{F} \to [-\infty, \infty]$ allowed to take at most one of the values in $\{-\infty, +\infty\}$ and satisfying:*

- $m(\emptyset) = 0$.
- *If $\{R_j\}_{j=1}^{\infty} \subseteq \mathcal{F}$ is a countable collection of disjoint sets, then $m\left(\cup_{j=1}^{\infty} R_j\right) = \sum_{j=1}^{\infty} m(R_j)$ as long as the series $\sum_{j=1}^{\infty} m(R_j)$ is absolutely convergent.*

Let $\{\Lambda_i\}_{i=1}^{q}$ be a finite set of probability measures on X with underlying σ-algebra $\mathcal{F}$ and $\{\alpha_i\}_{i=1}^{q}$ be (possibly negative) reals. Then $\Lambda = \sum_{i=1}^{q} \alpha_i \Lambda_i$ is a signed measure. Formally, for any $A \in \mathcal{F}$, $\Lambda(A) = \sum_{i=1}^{q} \alpha_i \Lambda_i(A)$. We will consider only such signed measures, arising as finite linear combinations of probability measures. Such a signed measure may identically vanish, i.e. $\Lambda(A) = 0 \; \forall A \in \mathcal{F}$. This is same as saying that if one writes $\Lambda = \Lambda' - \Lambda''$ as a difference of two non-negative measures (by grouping all Λ_i with positive and negative coefficients respectively), then Λ' and Λ'' are identical.

Gaussian Measures

Let Σ be an invertible, symmetric $t \times t$ matrix and μ be a t-dimensional vector. The Gaussian measure of a (measurable) set $A \subseteq \mathbb{R}^t$ w.r.t. *means* $\mu = (\mu_1, \ldots, \mu_t)$ and the *covariance matrix* Σ is defined as

$$\int_A \gamma_t(y = (y_1, \ldots, y_t), (\Sigma, \mu)) \;\; dy_1 dy_2 \ldots dy_t,$$

where $\gamma_t(\cdot, (\Sigma, \mu))$ is the Gaussian density function

$$\gamma_t(y, (\Sigma, \mu)) = \frac{1}{\sqrt{(2\pi)^t \mathrm{Det}(\Sigma)}} \; e^{-\frac{1}{2} \cdot (y-\mu)^T \Sigma^{-1} (y-\mu)}.$$

The random variables $y_1, \ldots, y_t$ then satisfy $\mathbb{E}[y_i] = \mu_i$ and $\Sigma_{ij} = \mathbb{E}[y_i y_j] - \mu_i \mu_j$.

With Σ, μ as above, one can also define a Gaussian measure on $(\mathbb{R}^d)^t$ that is a product measure with the measure on each of the d co-ordinates as above. Formally, if $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_t)$ with $\mathbf{y}_i \in \mathbb{R}^d$ and one denotes $\mathbf{y}^{(\ell)} \in \mathbb{R}^t$ as the vector of ℓ^{th} co-ordinates of $\mathbf{y}_1, \ldots, \mathbf{y}_t$ respectively for $\ell \in [d]$, then the measure is given by a density $\gamma_{t,d}(\mathbf{y}, (\Sigma, \mu))$ defined as:

$$\gamma_{t,d}(\mathbf{y}, (\Sigma, \mu)) = \prod_{\ell=1}^{d} \frac{1}{\sqrt{(2\pi)^t \mathrm{Det}(\Sigma)}} \; e^{-\frac{1}{2} \cdot (\mathbf{y}^{(\ell)} - \mu)^T \Sigma^{-1} (\mathbf{y}^{(\ell)} - \mu)}.$$

2.4 Our Results

In this section we present formal statements of our results. Given a predicate $f : \{-1, 1\}^k \to \{0, 1\}$, let $\mathcal{D}(f)$ denote the set of all probability distributions over $f^{-1}(1)$.

Definition 2.11 *For $\nu \in \mathcal{D}(f)$, we let $\zeta(\nu)$ denote the $(k+1) \times (k+1)$ symmetric* moment matrix *for ν such that:*

$$\begin{aligned} \forall i \in \{0\} \cup [k] : \; \zeta(i,i) &= 1 \,, \\ \forall i \in [k] : \; \zeta(0,i) &= \mathop{\mathbb{E}}_{x \sim \nu} [x_i] \,, \\ \forall i, j \in [k], i \neq j : \; \zeta(i,j) &= \mathop{\mathbb{E}}_{x \sim \nu} [x_i x_j] \,. \end{aligned}$$

Also, let $\mathcal{C}(f) \subseteq \mathbb{R}^{(k+1)\times(k+1)}$ denote the compact, convex set of all moment matrices:

$$\mathcal{C}(f) \;:=\; \{\zeta(\nu) \;:\; \nu \in \mathcal{D}(f)\}.$$

Note that the definition of the polytope $\mathcal{C}(f)$ defers somewhat from that in the introduction of the paper (it is now a $(k+1) \times (k+1)$ matrix as opposed to $\left(k + \binom{k}{2}\right)$-dimensional vector), but this difference is inconsequential.

For $S \subseteq [k]$, let ζ_S denote ζ restricted to the rows and columns of indices in $S \cup \{0\}$. For a permutation $\pi : S \to S$ we use $\zeta_{S,\pi}$ to denote a permutation of the submatrix ζ_S with the coordinates of S permuted according to π. Also, for a $|S|$-dimensional vector of signs $b \in \{-1, 1\}^{|S|}$, let $\zeta_{S,\pi,b} = \zeta_{S,\pi} \circ ((1 \; b)(1 \; b)^T)$ i.e., the matrix obtained by taking the Hadamard product (entrywise product) of the matrices $\zeta_{S,\pi}$ and $(1 \; b)(1 \; b)^T$.

Definition 2.12 *Let Λ be a probability measure supported on $\mathcal{C}(f)$. Then, for $S \subseteq [k]$, let Λ_S denote the measure on $(|S|+1) \times (|S|+1)$ matrices obtained by sampling $\zeta \sim \Lambda$ and taking the matrix ζ_S. Let $\pi : S \to S$ be any permutation and let $b \in \{-1,1\}^{|S|}$ be a vector of signs. We denote by $\Lambda_{S,\pi,b}$ the measure on $(|S|+1) \times (|S|+1)$ matrices obtained by sampling $\zeta \sim \Lambda$ and taking the matrix $\zeta_{S,\pi,b}$.*

We define a generic family of algorithms based on d-dimensional rounding of the vector solution to the basic relaxation, Figure 3. We choose to state an informal definition here as the exact rounding process is a but cumbersome, formally described in Subsection 3.2.

Definition 2.13 *(Informal) A d-dimensional rounding algorithm is a polynomial time algorithm based on an odd measurable function $\psi : \mathbb{R}^d \to [-1,1]$. The algorithm solves the basic relaxation for $\mathsf{CSP}(f)$, projects the SDP vectors onto a random d-dimensional subspace and then rounds them to $\{-1,1\}$ values according to (biases given by) ψ. The algorithm may draw the function ψ itself from a certain (pre-determined) distribution.*

Our main result appears below. It states that a predicate either admits a weak approximation based on a $(k+1)$-dimensional rounding algorithm or is strongly approximation resistant. This "dichotomy" is characterized precisely by the existence of a measure Λ on $\mathcal{C}(f)$ as in Definition 1.1.

Theorem 2.14 *Given $f : \{-1,1\}^k \to \{0,1\}$, the following "dichotomy" holds:*

- *Either there is a constant $\varepsilon > 0$ and a $(k+1)$-dimensional rounding algorithm that given a $(1-\varepsilon)$-satisfiable instance of $\mathsf{CSP}(f)$, outputs an assignment A such that $\mathbb{E}_A\left[|\mathsf{sat}(A) - \rho(f)|\right] \geq \varepsilon$ (i.e. achieves a weak approximation),*
- *Or there exists a probability measure Λ on $\mathcal{C}(f)$, such that for all $t \in [k]$, and a uniformly random choice of S with $|S| = t$, $\pi : S \to S$ and $b \in \{\pm 1\}^{|S|}$, the following signed measure on $(t+1) \times (t+1)$ matrices:*

$$\Lambda^{(t)} \;:=\; \mathop{\mathbb{E}}_{|S|=t} \; \mathop{\mathbb{E}}_{\pi:S\to S} \; \mathop{\mathbb{E}}_{b\in\{-1,1\}^{|S|}} \left[\left(\prod_{i \in S} b_i \right) \cdot \hat{f}(S) \cdot \Lambda_{S,\pi,b} \right] \tag{2.2}$$

is identically zero. In this case for every $\varepsilon > 0$, the predicate has a strong $(1-\varepsilon, \rho(f) \pm \varepsilon)$ integrality gap for the basic relaxation and (hence) for the mixed relaxation with a super-constant number of rounds and is strongly approximation resistant, i.e. $\mathsf{GapCSP}(f)_{1-\varepsilon,\rho(f)\pm\varepsilon}$ is UG-hard.

Similarly, we obtain a "dichotomy" for the integrality gap in the Sherali-Adams LP hierarchy. The characterization is syntactically similar once the polytope $\mathcal{C}(f)$ is replaced by the polytope $\mathcal{C}^*(f)$ consisting of only the first moment vectors of distributions supported on $f^{-1}(1)$ (and is therefore the same as the convex hull of $f^{-1}(1)$). For a measure Λ^* on $\mathcal{C}^*(f)$ and a subset $S \subseteq [k]$, the projected measure Λ^*_S and the measure $\Lambda^*_{S,\pi,b}$ for a permutation $\pi : S \mapsto S$ and signs $b \in \{-1,1\}^S$ are defined in an analogous manner. The family of generic algorithms is now defined w.r.t. only the first moments, i.e. the algorithm can "use" only the biases computed by the LP relaxation.

Definition 2.15 *For $\nu \in \mathcal{D}(f)$, we let $\zeta(\nu)$ denote the k-dimensional vector such that:*

$$\forall i \in [k] : \; \zeta(i) \;:=\; \mathop{\mathbb{E}}_{x\sim\nu}[x_i] \,.$$

Let $\mathcal{C}^*(f) \subseteq \mathbb{R}^k$ denote the convex, compact set:

$$\mathcal{C}^*(f) \;:=\; \{\zeta(\nu) : \nu \in \mathcal{D}(f)\}.$$

Definition 2.16 *(Informal) A k-round* LP rounding algorithm *is a polynomial time algorithm based on an odd measurable function $\psi : [-1,1] \to [-1,1]$. The algorithm solves the k-round Sherali-Adams relaxation for $\mathsf{CSP}(f)$ and then a CSP variable with bias p (as computed by the LP relaxation) is rounded to a $\{-1,1\}$ value with bias $\psi(p)$, independently for different variables. The algorithm may draw the function ψ itself from a certain (pre-determined) distribution.*

Theorem 2.17 *Given $f : \{-1,1\}^k \to \{0,1\}$, the following "dichotomy" holds:*

- *Either there is a constant $\varepsilon > 0$ and a k-round LP rounding algorithm that given a $(1-\varepsilon)$-satisfiable instance of $\mathsf{CSP}(f)$, outputs an assignment A such that $\mathbb{E}_A[|\mathsf{sat}(A) - \rho(f)|] \geq \varepsilon$ (i.e. achieves a weak approximation),*
- *Or there exists a probability measure Λ^* on $\mathcal{C}^*(f)$, such that for all $t \in [k]$, and a uniformly random choice of S with $|S| = t$, $\pi : S \to S$ and $b \in \{\pm 1\}^{|S|}$, the following signed measure on t-dimensional vectors:*

$$\Lambda^{*,(t)} \;:=\; \mathop{\mathbb{E}}_{|S|=t} \; \mathop{\mathbb{E}}_{\pi:S\to S} \; \mathop{\mathbb{E}}_{b\in\{-1,1\}^{|S|}} \left[\left(\prod_{i\in S} b_i \right) \cdot \hat{f}(S) \cdot \Lambda^*_{S,\pi,b} \right] \tag{2.3}$$

  is identically zero. In this case for every $\varepsilon > 0$, the predicate has a strong $(1-\varepsilon, \rho(f) \pm \varepsilon)$ integrality gap for a super-constant number of rounds of the Sherali-Adams LP relaxation.

As described in the introduction, we obtain additional interesting observations and results (e.g. approximation resistance for odd predicates and k-partite version of CSPs). We skip their formal statements and proofs from the current version of the paper. The proofs of our main results, namely Theorem 2.14 and Theorem 2.17, appear in Section 3 and Section 4 respectively.

2.5 The Unique Games Conjecture

We present the definitions of the Unique Games problem and UG-hardness.

Definition 2.18 *A Unique Games instance $\mathcal{L}(G(V,E),[L],\{\pi_{v,w}\}_{(v,w)\in E})$ consists of a graph $G(V,E)$, a set of labels $[L]$ and a set of permutations $\pi_{v,w} : [L] \mapsto [L]$, one for each edge of the graph (the edges have an implicit direction). A labeling is an assignment $A : V \mapsto [L]$. The labeling satisfies an edge (v,w) if $\pi_{v,w}(A(v)) = A(w)$. $\mathsf{OPT}(\mathcal{L})$ is the maximum fraction of edges satisfied by any labeling.*

Let $\mathsf{GapUG}_{1-\delta,\delta}$ denote the gap version of the Unique Games problem where the instance $\mathcal{L}$ is promised to have either $\mathsf{OPT}(\mathcal{L}) \geq 1-\delta$ or $\mathsf{OPT}(\mathcal{L}) \leq \delta$. Khot [24] conjectures that for an arbitrarily small constant $\delta > 0$, $\mathsf{GapUG}_{1-\delta,\delta}$ is NP-hard on instances with L labels where $L = L(\delta)$ may depend on δ.

Definition 2.19 *A decision problem is said to be UG-hard if for a sufficiently small constant $\delta > 0$, there is a polynomial time reduction from $\mathsf{GapUG}_{1-\delta,\delta}$ (with the number of labels $L = L(\delta)$) to the problem under consideration.*

3 Proof of the SDP Dichotomy Theorem

In this section we present the proof of Theorem 2.14. We begin by developing the game theory formalism that we will need. This formalism is mostly common to both dichotomy theorems but we state it first with the SDPs in mind.

3.1 Game-Theoretic Formulation

We have two players: Alice, the player trying to design an algorithm, and Harry, the player trying to prove a hardness result. Intuitively, Alice wants to show that any $\mathsf{CSP}(f)$ admits a non-trivial efficient approximation via rounding the natural SDP relaxation. For this, Alice will try to maximize the pay-off in the (zero-sum) game, which we shall define soon. On the other hand, Harry intends to minimize the pay-off. Intuitively, Harry wants to show that there exist $\mathsf{CSP}(f)$ instances for which the integrality gap is high.

The *pure strategies* of Harry will correspond to *distributions* over moment matrices. Recall that $\mathcal{C}(f)$ was the set of $(k+1) \times (k+1)$ moment matrices for distributions in $\mathcal{D}(f)$. To ensure that our moment matrices are non-singular, we will actually need to work with a slightly modified body $\mathcal{C}_\delta(f)$ defined as

$$\mathcal{C}_\delta(f) \ := \ \{(1-\delta)\cdot\zeta + \delta\cdot\mathbb{I}_{k+1} \ : \ \zeta \in \mathcal{C}(f)\} \ ,$$

for a constant $\delta \in (0,1)$. Here, $\mathbb{I}_{k+1}$ denotes the $(k+1)\times(k+1)$ identity matrix. Let R_i, such that $R_1 \subseteq R_2 \ldots \subseteq R_p \subseteq \ldots$ denote a fixed sequence of finite subsets of $\mathcal{C}_\delta(f)$. We assume that the above sequence is dense in $\mathcal{C}_\delta(f)$ in the limit. Let $\mathcal{R}_p$ denote the class of distributions over R_p, such that all the probabilities are integral multiples of $1/2^p$.

Let $\{\mathcal{P}_q\}_{q\in\mathbb{N}}$ denote a sequence of partitions, where $\mathcal{P}_q$ partitions $[-1,1]^d$ into $2^{(q+1)d}$ boxes of equal size. We will choose $d = k+1$ for reasons that will become clear later. Note that for each q, $\mathcal{P}_{q+1}$ is a refinement of $\mathcal{P}_q$ inside. Let V_k denote the set of values $\left\{\frac{r}{k} \ : \ r \in \mathbb{Z}, -k \le r \le k\right\}$ and let $\psi_q : \mathbb{R}^d \to V_k$ denote an *odd* function which takes values in V_k in $[-1,1]^d$ and 0 outside. We further assume that ψ_q is constant on each cell of $\mathcal{P}_q$. This can be ensured since the partitions $\mathcal{P}_q$ are symmetric with respect to 0.

For every fixed p, q above, we will define a zero-sum game $\mathcal{G}_{p,q}$. A pure strategy for Harry corresponds to a distribution $\lambda \in \mathcal{R}_p$, and a mixed strategy is a probability distribution Λ_p over $\mathcal{R}_p$. The pure strategies for Alice are given by all possible functions ψ_q, and a mixed strategy is a probability distribution Γ_q over these. For $\lambda \in \mathcal{R}_p$ and ψ defined over $\mathbb{R}^d$ as described above, the payoff of the (pure strategy) game $\mathcal{G}_{p,q}$ is given by:

$$\mathsf{PayOff}(\lambda,\psi) \ := \ \left| \mathop{\mathbb{E}}_{\zeta\sim\lambda} \ \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} \left[\sum_{S\neq\emptyset} \hat{f}(S)\cdot\prod_{i\in S}\psi(\mathbf{y}_i) \right] \right| , \tag{3.1}$$

where $\{\mathbf{y}_i\}_{i\in[k]}$ are points in $\mathbb{R}^d$ sampled from a Gaussian process $\mathcal{N}_d(\zeta)$ different coordinates being independent and for each coordinate $l \in [d]$, $\mathbb{E}\left[(\mathbf{y}_i)_l\right] = \zeta(0,i)$ and $\mathbb{E}\left[(\mathbf{y}_i)_l(\mathbf{y}_j)_l\right] = \zeta(i,j)\ \forall i,j \in [k]$. Note that this corresponds to the deviation from a random assignment that would be obtained by the rounding algorithm corresponding to ψ, if the SDP vectors corresponding to variables in each constraint had correlations given by ζ. We shall also use the notation $\mathsf{PayOff}(\lambda,\psi)$ to denote the above expression, for an arbitrary distribution λ over $\mathcal{C}_\delta(f)$ and any function $\psi : \mathbb{R}^d \to V_k$. For a matrix ζ, we use $\mathsf{PayOff}(\zeta,\psi)$ to denote $\mathsf{PayOff}(\lambda,\psi)$ for a distribution λ concentrated on a single point corresponding to ζ.

For mixed strategies Λ_p and Γ_q the payoff is given by:

$$\mathsf{PayOff}(\Lambda_p, \Gamma_q) \;:=\; \mathop{\mathbb{E}}_{\lambda\sim\Lambda_p} \; \mathop{\mathbb{E}}_{\psi\sim\Gamma_q} \left| \mathop{\mathbb{E}}_{\zeta\sim\lambda} \; \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} \left[\sum_{S\neq\emptyset} \hat{f}(S) \cdot \prod_{i\in S} \psi(\mathbf{y}_i) \right] \right| . \tag{3.2}$$

Alice plays to maximize the above payoff and Harry plays to minimize the same. By Von Neumann's min-max theorem there exists a unique value for the above game $\mathcal{G}_{p,q}$, for every p and q. Our next task will be to relate the value of this game (in the limit) to the hardness of the predicate f.

The following simple fact about the pay-off function will be used repeatedly.

Claim 3.1 *For any λ and $\psi : \mathbb{R}^d \to [-1,1]$, $\mathsf{PayOff}(\lambda,\psi) \leq 1$.*

Proof: Since the pay-off function is

$$\mathsf{PayOff}(\lambda, \psi) \;:=\; \left| \mathop{\mathbb{E}}_{\zeta\sim\lambda} \; \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} \left[\sum_{S\neq\emptyset} \hat{f}(S) \cdot \prod_{i\in S} \psi(\mathbf{y}_i) \right] \right| ,$$

it suffices to show that for any k values $\psi(\mathbf{y}_1), \ldots, \psi(\mathbf{y}_k) \in V_k$, we have $\left|\sum_{S\neq\emptyset} \hat{f}(S) \cdot \prod_{i\in S} \psi(\mathbf{y}_i)\right| \leq 1$. This is immediate if all the values are ± 1, since we can write the expression inside the absolute value as $f(\psi(\mathbf{y}_1), \ldots, \psi(\mathbf{y}_k)) - \hat{f}(\emptyset)$, which only takes values $-\hat{f}(\emptyset)$ and $1-\hat{f}(\emptyset)$ since f is Boolean. If a value, say $\psi(\mathbf{y}_i)$ is in $(-1,1)$, we can view the expression as an expectation over the cases where we choose the i^{th} input variable of f to be 1 with probability $(1+\psi(\mathbf{y}_i))/2$ and -1 with probability $(1-\psi(\mathbf{y}_i)/2)$, which shows that it is always between $-\hat{f}(\emptyset)$ and $1-\hat{f}(\emptyset)$. ∎

We will also need another simple fact about a matrix $\zeta \in \mathcal{C}_\delta(f)$. Recall that for a matrix ζ, for $S \subseteq [k]$, $\pi : S \to S$ and $b \in \{-1,1\}^S$, we define the matrix $\zeta_{S,\pi,b}$ by considering the submatrix given by rows and columns in S (and the first row and first column), permuting them according to π and multiplying each row i by b_i and each column j by b_j (thus, the (i,j) entry is multiplied by $b_i \cdot b_j$). Also, for each matrix $\zeta' = \zeta_{S,\pi,b}$, we can define a *covariance matrix* Σ, with $\Sigma_{ij} = \zeta'(i,j) - \zeta'(0,i)\cdot\zeta'(0,j)$. Then we shall use the following fact repeatedly.

Claim 3.2 *Let $\zeta \in \mathcal{C}_\delta(f)$ and let $\zeta_{S,\pi,b}$ be as defined above for an arbitrary choice of S, π and b. Let Σ be the covariance matrix corresponding to $\zeta_{S,\pi,b}$. Then Σ is a positive semidefinite matrix with all eigenvalues at least δ.*

Proof: Note that it is sufficient to prove the claim with π being the identity permutation and $b = 1^k$, since permuting the rows and columns, or multiplying them with a sign does not affect the eigenvalues. Let us first consider the case when $S = [k]$ and $b = 1^k$ (and thus $\zeta_{S,\pi,b} = \zeta$). For $\zeta \in \mathcal{C}_\delta(f)$, there is a distribution ν on $f^{-1}(1)$, and in particular on $\{-1,1\}^k$, so that $\zeta = (1-\delta)\cdot\zeta(\nu) + \delta\cdot\mathbb{I}_{k+1}$. Let $\tilde{\zeta}$ denote $\zeta(\nu)$. If $\Sigma(\zeta)$ denotes the covariance matrix corresponding to ζ, then we can write

$$\Sigma(\zeta) \;=\; (1-\delta)\cdot\Sigma(\tilde{\zeta}) + \delta\cdot(1-\delta)\cdot M + \delta\cdot\mathbb{I}_k\,,$$

where M is a positive semidefinite (PSD) matrix with $M_{ij} = \tilde{\zeta}(0,i)\cdot\tilde{\zeta}(0,j)$. Also, note that $\Sigma(\tilde{\zeta})$ is a covariance matrix corresponding to a distribution ν on $\{-1,1\}^k$ and is hence PSD. Thus, all eigenvalues for $\Sigma(\zeta)$ are at least δ. Similarly, when $|S| = t$ for some $t \leq k$, we can consider ν_S, the projection of ν to $\{-1,1\}^S$. We can again write $\zeta_S = (1-\delta)\cdot\zeta(\nu_S) + \delta\cdot\mathbb{I}_{|S|+1}$. The rest of the proof is same as above. ∎

Let $\mathcal{V}$ denote the (infinite) matrix over reals such that: $\mathcal{V}(p,q) := \mathsf{val}(\mathcal{G}_{p,q})$ (the equilibrium value for the game $\mathcal{G}_{p,q}$).

Lemma 3.3 *The limit L defined below exists and it is finite.*

$$L \;:=\; \lim_{p,q\to\infty} \mathcal{V}(p,q)\,. \tag{3.3}$$

Moreover, every row p has a limit r_p as $q \to \infty$ and every column q has a limit c_q as $p \to \infty$.

Proof: First, observe that the entries in $\mathcal{V}$ are all non-negative and bounded above by 1 (using Claim 3.1). Also, for any fixed q, $\mathcal{V}(p,q)$ is non-increasing as p increases since $\mathcal{R}_p \subseteq \mathcal{R}_{p+1}$. Similarly, for any fixed p, $\mathcal{V}(p,q)$ is non-decreasing as q increases. This follows from the fact that $\mathcal{P}_{q+1}$ is a refinement of $\mathcal{P}_q$ and thus each strategy ψ_q can also be implemented by a function ψ_{q+1}.

Therefore, by the monotone convergence theorem every row (resp. column) in $\mathcal{G}$ has a limit, say r_p (resp. c_q). Moreover, r_p is non-increasing as p increases and c_q is non-decreasing as q increases. Therefore, again by the monotone convergence theorem both these sequences have to converge. Also, the limits must coincide since for sufficiently large p, q, $\mathcal{V}(p,q)$ must come arbitrarily close to both the limits. This common limit is denoted by L in our statement. ∎

We shall also need the following lemma.

Lemma 3.4 *If $\zeta, \zeta' \in \mathcal{C}_\delta(f)$ are such that $\|\zeta - \zeta'\|_\infty \le \varepsilon$, then for any function $\psi : \mathbb{R}^d \to V_k$, we have*

$$\left| \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} \left[\sum_{S\neq\emptyset} \hat{f}(S)\cdot\prod_{i\in S}\psi(\mathbf{y}_i) \right] - \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta')} \left[\sum_{S\neq\emptyset} \hat{f}(S)\cdot\prod_{i\in S}\psi(\mathbf{y}_i) \right] \right| \;=\; O_{k,d,\delta}(\varepsilon)\,.$$

Hence, the function $\mathsf{PayOff}(\zeta,\psi)$ is $O_{k,d,\delta}(1)$-Lipschitz in the argument ζ.

Proof: Let $g(\mathbf{y}_1,\ldots,\mathbf{y}_k)$ denote the expression $\sum_{S\neq\emptyset}\hat{f}(S)\cdot\prod_{i\in S}\psi(\mathbf{y}_i)$. We first note that proving the bound claimed above is also sufficient to show that $\mathsf{PayOff}(\zeta,\psi)$ is $O_{k,d,\delta}(1)$-Lipschitz in the argument ζ, since $\mathsf{PayOff}(\zeta,\psi) = \left|\mathbb{E}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)}\left[g(\mathbf{y}_1,\ldots,\mathbf{y}_k)\right]\right|$ and hence,

$$\left|\mathsf{PayOff}(\zeta,\psi) - \mathsf{PayOff}(\zeta',\psi)\right| \;\le\; \left| \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} [g(\mathbf{y}_1,\ldots,\mathbf{y}_k)] - \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta')} [g(\mathbf{y}_1,\ldots,\mathbf{y}_k)] \right|.$$

By Claim 3.1, we have that $|g(\mathbf{y}_1,\ldots,\mathbf{y}_k)| \le 1$. Hence, we can bound the above expression as

$$\left| \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} [g(\mathbf{y}_1,\ldots,\mathbf{y}_k)] - \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta')} [g(\mathbf{y}_1,\ldots,\mathbf{y}_k)] \right| \;\le\; 2\cdot\left\|\mathcal{N}_d(\zeta) - \mathcal{N}_d(\zeta')\right\|_1\,,$$

where $\|\mathcal{N}_d(\zeta) - \mathcal{N}_d(\zeta')\|_1$ denotes the total variation distance between the two distributions. This can be bounded by $O_{k,d,\delta}(\varepsilon)$ as below.

By Pinsker's inequality one can bound the total variation distance by the Kullback-Leibler (KL) divergence, denoted $D(\mathcal{N}_d(\zeta)||\mathcal{N}_d(\zeta'))$, as follows.

$$\begin{aligned}\left\|\mathcal{N}_d(\zeta) - \mathcal{N}_d(\zeta')\right\|_1^2 \;&\le\; \frac{1}{2}\cdot\left[D(\mathcal{N}_d(\zeta)\|\mathcal{N}_d(\zeta')) + D(\mathcal{N}_d(\zeta')\|\mathcal{N}_d(\zeta))\right] \\ &=\; \frac{d}{2}\cdot\left[D(\mathcal{N}(\zeta)\|\mathcal{N}(\zeta')) + D(\mathcal{N}(\zeta')\|\mathcal{N}(\zeta))\right]\,,\end{aligned}$$

where the equality uses the fact that $\mathcal{N}_d(\zeta)$ and $\mathcal{N}_d(\zeta')$ are product distributions of d k-dimensional Gaussians and $D((P_1,P_2)\|(Q_1,Q_2)) = D(P_1\|Q_1) + D(P_2\|Q_2)$ for product distributions (P_1,P_2) and (Q_1,Q_2). The sum on the right can now be bounded by $O_{k,d,\delta}(\varepsilon^2)$.

Let Σ and Σ' denote the covariance matrices for ζ and ζ'. Let μ and μ' denote the vector of means for ζ and ζ'. For a multivariate normal distribution in k dimensions the sum of the two KL divergences, as above, can be written as follows (for eg. see chapter 15 in [22])

$$D(\mathcal{N}(\zeta)\|\mathcal{N}(\zeta'))+D(\mathcal{N}(\zeta')\|\mathcal{N}(\zeta)) \;=\; (\mu-\mu')^T\left(\frac{\Sigma^{-1}+\Sigma'^{-1}}{2}\right)(\mu-\mu') - \frac{1}{2}\cdot(\Sigma-\Sigma')\bullet(\Sigma^{-1}-\Sigma'^{-1})\,.$$

Since $\|\mu-\mu'\|_\infty \le \|\zeta-\zeta'\|_\infty \le \varepsilon$ and all eigenvalues of Σ^{-1} and Σ'^{-1} are at most $1/\delta$, the first term is bounded by $O_{k,\delta}(\varepsilon^2)$. For bounding Frobenius product in the second term, note that $\|\Sigma-\Sigma'\|_\infty = O(\varepsilon)$. For the term, $\Sigma^{-1}-\Sigma'^{-1}$, using the fact that $\Sigma^{-1} = \frac{\mathrm{Adj}(\Sigma)}{|\Sigma|}$, we can write

$$\Sigma^{-1}-\Sigma'^{-1} \;=\; \frac{\mathrm{Adj}(\Sigma)\cdot|\Sigma'| - \mathrm{Adj}(\Sigma)\cdot|\Sigma'|}{|\Sigma|\cdot|\Sigma'|}\,.$$

Since all eigenvalues of Σ and Σ' are at least δ, the determinants $|\Sigma|$ and $|\Sigma'|$ are at least δ^k. Each entry of the matrices in the numerator can be viewed as a difference between two multivariate degree-$2k$ polynomials with $O_k(1)$ terms. The two polynomials are identical, except that each has been perturbed by at most ε in its variables. Hence, their difference can be at most $O_k(\varepsilon)$.

Thus, we obtain that $\left\|\Sigma^{-1}-\Sigma'^{-1}\right\|_\infty = O_{k,\delta}(\varepsilon)$ and hence $(\Sigma-\Sigma')\bullet(\Sigma^{-1}-\Sigma'^{-1}) = O_{k,\delta}(\varepsilon^2)$, which gives the required bound on $\|\mathcal{N}_d(\zeta)-\mathcal{N}_d(\zeta')\|_1$. ∎

3.2 A Rounding Scheme for Predicates when $L > 0$

We can now prove that if the value of the above games has a positive limit, then the predicate f admits a non-trivial weak approximation. For an instance Φ of $\mathsf{CSP}(f)$ and a d-dimensional rounding function $\psi : \mathbb{R}^d \to V_k$, let $\mathsf{ROUND}_\psi(\Phi)$ denote the expected fraction of constraints satisfied by the rounding algorithm using the function ψ. Recall that a d-dimensional *rounding algorithm* is a distribution over functions ψ. We will show that if the value of the game is positive, then there is a rounding scheme such that $\mathbb{E}_\psi\,|\mathsf{ROUND}_\psi - \rho(f)|$ is $\Omega(1)$.

Theorem 3.5 *If $L > 0$, then there exists a $(k+1)$-dimensional rounding algorithm for the basic SDP relaxation of $\mathsf{CSP}(f)$, such that given an instance Φ with $\mathsf{FRAC}(\Phi) \ge 1-\varepsilon$ (for sufficiently small $\varepsilon > 0$), we have $\mathbb{E}_\psi\,|\mathsf{ROUND}_\psi(\Phi) - \rho(f)| \ge L/2$.*

Proof: Since $\lim_{p,q\to\infty}\mathsf{val}(\mathcal{G}_{p,q}) = L > 0$, for every $\beta > 0$ there exists a $q \in \mathbb{N}$, such that for all $p \in \mathbb{N}$, $\mathsf{val}(\mathcal{G}_{p,q}) \ge L-\beta$. This means that for all $p \in \mathbb{N}$, there exists a distribution Γ_q over functions ψ_q, such that for all $\lambda \in \mathcal{R}_p$

$$\mathsf{PayOff}(\lambda,\Gamma_q) \;\ge\; L-\beta\,.$$

We will fix a sufficiently large p later. We will use the corresponding Γ_q to design a d-dimensional rounding strategy (recall that we choose $d = k+1$). Given an instance Φ of $\mathsf{CSP}(f)$ and a solution to the SDP in Figure 3, we proceed as follows:

- For all $i \in [n]$, define vectors:

$$\begin{aligned} \mathbf{u}_\emptyset &= \mathbf{v}_{(\emptyset,\emptyset)} \\ \tilde{\mathbf{u}}_i &= \mathbf{v}_{(i,1)} - \mathbf{v}_{(i,-1)} \\ \mathbf{u}_i &= \sqrt{1-\delta} \cdot \tilde{\mathbf{u}}_i + \sqrt{\delta} \cdot \mathbf{e}_i \,, \end{aligned}$$

  where $\{\mathbf{e}_i\}_{i\in[n]}$ form an orthonormal basis, orthogonal to all the vectors $\left\{\mathbf{v}_{(i,b)}\right\}_{i\in[n],b\in\{-1,1\}}$.

- Sample vectors $\{\mathbf{g}_l\}_{l\in[d]}$ such that each coordinate of each $\mathbf{g}_l$ is a standard normal variable. Define the vectors $\mathbf{y}'_1, \ldots, \mathbf{y}'_n \in \mathbb{R}^d$ such that for each $l \in [d]$,

$$(\mathbf{y}'_i)_l \;=\; \langle (\mathbf{u}_i - \langle \mathbf{u}_i, \mathbf{u}_\emptyset \rangle \cdot \mathbf{u}_\emptyset)\,, \mathbf{g}_l \rangle + \langle \mathbf{u}_i, \mathbf{u}_\emptyset \rangle$$

- Sample $\psi \sim \Gamma_q$. For each $i \in [n]$, assign the variable x_i as 1 with probability $(1+\psi(\mathbf{y}'_i))/2$ and -1 with probability $(1-\psi(\mathbf{y}'_i))/2$.

For a constraint C, let $\mathsf{FRAC}(C) \in [0,1]$ denote the contribution of the constraint C to the SDP objective function. For the given instance Φ, we have $\mathsf{FRAC}(\Phi) = \mathbb{E}_{C\in\Phi}\left[\mathsf{FRAC}(C)\right] \geq 1-\varepsilon$ and hence $\mathbb{P}_{C\in\Phi}\left[\mathsf{FRAC}(C) \geq 1-\sqrt{\varepsilon}\right] \geq 1-\sqrt{\varepsilon}$. Let C be a constraint such that $\mathsf{FRAC}(C) \geq 1-\sqrt{\varepsilon}$. Without loss of generality, we can take C to be on the variables $x_1, \ldots, x_k$ and of the form $f(x_1 \cdot b_1, \ldots, x_k \cdot b_k)$ for $b_1, \ldots, b_k \in \{-1, 1\}$.

The probability that C is satisfied by the assignment produced by a rounding function ψ, chosen by our rounding scheme is given by

$$\begin{aligned} \mathsf{ROUND}_\psi(C) &= \rho(f) + \mathop{\mathbb{E}}_{\mathbf{y}'_1,\ldots,\mathbf{y}'_k} \left[\sum_{\substack{S\subseteq[k] \\ S\neq\emptyset}} \left(\prod_{i\in S} b_i \right) \cdot \hat{f}(S) \cdot \left(\prod_{i\in S} \psi(\mathbf{y}'_i) \right) \right] \\ &= \rho(f) + \mathop{\mathbb{E}}_{\mathbf{y}'_1,\ldots,\mathbf{y}'_k} \left[\sum_{\substack{S\subseteq[k] \\ S\neq\emptyset}} \hat{f}(S) \cdot \left(\prod_{i\in S} \psi(b_i \cdot \mathbf{y}'_i) \right) \right] , \end{aligned}$$

where $b_i \cdot \mathbf{y}'_i$ denotes a vector with each coordinate multiplied by b_i, and the second equality used the fact that the functions ψ are odd.

Let $\zeta_C \in \mathbb{R}^{(k+1)\times(k+1)}$ be the symmetric moment matrix with $\zeta(0,i) = \langle b_i \cdot \mathbf{u}_i, \mathbf{u}_\emptyset \rangle$ and $\zeta(i,j) = \langle b_i \cdot \mathbf{u}_i, b_j \cdot \mathbf{u}_j \rangle$. Then the variables $(\mathbf{y}_1, \ldots, \mathbf{y}_k) = (b_1 \cdot \mathbf{y}'_1, \ldots, b_k \cdot \mathbf{y}'_k)$ are distributed according to the Gaussian process $\mathcal{N}_d(\zeta_C)$. Thus, we can write

$$\mathsf{ROUND}_\psi(C) \;=\; \rho(f) + \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta_C)} \left[\sum_{\substack{S\subseteq[k] \\ S\neq\emptyset}} \hat{f}(S) \cdot \left(\prod_{i\in S} \psi(\mathbf{y}_i) \right) \right] .$$

The variables $x_{([k],\alpha)}$ define a probability distribution, say ν_0 on $\{-1,1\}^k$. Let ν be the distribution on $\{-1,1\}^k$ such that for any $x \in \{-1,1\}^k$,

$$\nu(x_1, \ldots, x_k) \;=\; \nu_0(b_1 \cdot x_1, \ldots, b_k \cdot x_k)\,.$$

Then $\mathbb{P}_{x\sim\nu}[f(x)=1] \geq 1-\sqrt{\varepsilon}$ and for the corresponding moment matrix $\zeta(\nu)$, we have $\zeta(\nu)(i,j) = \langle b_i \cdot \tilde{\mathbf{u}}_i, b_j \cdot \tilde{\mathbf{u}}_j \rangle$ and $\zeta(\nu)(0,i) = \langle b_i \cdot \tilde{\mathbf{u}}_i, \mathbf{u}_\emptyset \rangle$ for all $i,j \in [k]$. From the definition of the vectors $\mathbf{u}_i$ and the matrix ζ_C above, we have that

$$\zeta_C \;=\; (1-\delta)\cdot\zeta(\nu) + \delta\cdot\mathbb{I}\,.$$

However, ζ_C does not lie in the body $\mathcal{C}_\delta(f)$ since ν is not entirely supported on $f^{-1}(1)$. We thus, consider the distribution ν', which is ν conditioned on the output being in $f^{-1}(1)$. Also, we define the matrix $\zeta'_C \in \mathcal{C}_\delta(f)$ as

$$\zeta'_C \;:=\; (1-\delta)\cdot\zeta(\nu') + \delta\cdot\mathbb{I}\,.$$

Since ν satisfies C with probability at least $1-\sqrt{\varepsilon}$, we have $\|\nu-\nu'\|_1 = O(\sqrt{\varepsilon})$. Also, this gives that $\|\zeta_C - \zeta'_C\|_\infty \leq \sqrt{\varepsilon}$. By Lemma 3.4, we have for ζ_C and ζ'_C as above

$$\left|\mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta_C)}\left[\sum_{\substack{S\subseteq[k]\\S\neq\emptyset}}\hat{f}(S)\cdot\left(\prod_{i\in S}\psi(\mathbf{y}_i)\right)\right] - \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta'_C)}\left[\sum_{\substack{S\subseteq[k]\\S\neq\emptyset}}\hat{f}(S)\cdot\left(\prod_{i\in S}\psi(\mathbf{y}_i)\right)\right]\right| = O_{k,d,\delta}(\sqrt{\varepsilon})\,.$$

We now analyze $\mathsf{ROUND}_\psi(\Phi) = \mathbb{E}_{C\in\Phi}[\mathsf{ROUND}_\psi(C)]$. Let Φ' denote the instance restricted to the constraints C such that $\mathsf{FRAC}(C) \geq 1-\sqrt{\varepsilon}$. Since $\mathbb{P}_{C\in\Phi}[\mathsf{FRAC}(C) \geq 1-\sqrt{\varepsilon}] \geq 1-\sqrt{\varepsilon}$, we have that

$$\left|\mathsf{ROUND}_\psi(\Phi) - \mathsf{ROUND}_\psi(\Phi')\right| \;=\; O(\sqrt{\varepsilon})\,.$$

Finally, to relate the above to the value of one of the games $\mathcal{G}_{p,q}$, we let λ be the distribution on $\mathcal{C}_\delta(f)$ obtained by sampling a random $C\in\Phi'$ and taking the matrix ζ'_C. Using the above, we get that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\psi\sim\Gamma_q}\left|\mathsf{ROUND}_\psi(\Phi) - \rho(f)\right| &\geq \mathop{\mathbb{E}}_{\psi\sim\Gamma_q}\left|\mathsf{ROUND}_\psi(\Phi') - \rho(f)\right| - O(\sqrt{\varepsilon})\\
&= \mathop{\mathbb{E}}_{\psi\sim\Gamma_q}\left|\mathop{\mathbb{E}}_{C\sim\Phi'}\;\mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta_C)}\left[\sum_{\substack{S\subseteq[k]\\S\neq\emptyset}}\hat{f}(S)\cdot\left(\prod_{i\in S}\psi(\mathbf{y}_i)\right)\right]\right| - O(\sqrt{\varepsilon})\\
&\geq \mathop{\mathbb{E}}_{\psi\sim\Gamma_q}\left|\mathop{\mathbb{E}}_{C\sim\Phi'}\;\mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta'_C)}\left[\sum_{\substack{S\subseteq[k]\\S\neq\emptyset}}\hat{f}(S)\cdot\left(\prod_{i\in S}\psi(\mathbf{y}_i)\right)\right]\right| - O_{k,d,\delta}(\sqrt{\varepsilon})\\
&= \mathop{\mathbb{E}}_{\psi\sim\Gamma_q}\left|\mathop{\mathbb{E}}_{\zeta\sim\lambda}\;\mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)}\left[\sum_{\substack{S\subseteq[k]\\S\neq\emptyset}}\hat{f}(S)\cdot\left(\prod_{i\in S}\psi(\mathbf{y}_i)\right)\right]\right| - O_{k,d,\delta}(\sqrt{\varepsilon})\\
&= \mathsf{PayOff}(\lambda,\Gamma_q) - O_{k,d,\delta}(\sqrt{\varepsilon})\,.
\end{aligned}$$

The above almost looks like the pay-off for our game, except for the fact that the distribution λ may not belong to the set of distributions $\mathcal{R}_p$ for any $p\in\mathbb{N}$. However, the sets R_p get arbitrarily dense in $\mathcal{C}_\delta(f)$ as p increases. Also the probabilities for distributions in $\mathcal{R}_p$ are allowed to be multiples of $1/2^p$, which gets arbitrarily small as p increases. Since λ is supported on finitely many $\zeta\in\mathcal{C}_\delta(f)$, for any $\beta_0>0$, it is possible to find a large enough $p\in\mathbb{N}$ and a distribution $\lambda_p\in\mathcal{R}_p$ such that:

- There exists a bijection between the supports of λ and λ_p.

- For each ζ in the support of λ, let ζ' denote its image according to the above map. Then $\|\zeta - \zeta'\|_\infty \le \beta_0$ and the probabilities $\lambda(\zeta)$ and $\lambda_p(\zeta')$ differ by at most β_0.

Also, since by Lemma 3.4 the function $\mathsf{PayOff}(\zeta, \psi_q)$ is $O_{k,d,\delta}(1)$-Lipschitz in the argument ζ, we have that there exists a $p \in \mathbb{N}$ and a distribution $\lambda_p \in \mathcal{R}_p$ such that

$$|\mathsf{PayOff}(\lambda, \Gamma_q) - \mathsf{PayOff}(\lambda_p, \Gamma_q)| \le \beta\,.$$

Finally, by definition of Γ_q, we have that

$$\mathsf{PayOff}(\lambda_p, \Gamma_q) \;\ge\; \mathsf{val}(\mathcal{G}_{p,q}) \;\ge\; L - \beta\,.$$

Combining all the above inequalities, we have that

$$\mathop{\mathbb{E}}_{\psi} |\mathsf{ROUND}_\psi(\Phi) - \rho(f)| \;\ge\; \mathsf{PayOff}(\lambda_p, \Gamma_q) - \beta - O_{k,d,\delta}(\sqrt{\varepsilon}) \;\ge\; L - 2\beta - O_{k,d,\delta}(\sqrt{\varepsilon})\,.$$

Choosing $\beta \le L/8$ and $\varepsilon = o_{k,d,\delta}(L^2)$ gives that $\mathbb{E}_\psi |\mathsf{ROUND}_\psi(\Phi) - \rho(f)| \ge L/2$ as claimed. ∎

3.3 A Characterization of Predicates with $L = 0$

We are now left with the case: $L = 0$ (since we always have that $L \ge 0$). We will show that the condition $L = 0$ implies the existence of a probability measure Λ on $\mathcal{C}(f)$ satisfying Equation 2.2 in Theorem 2.14. We next prove the following.

Theorem 3.6 *If $L = 0$, then there exists a probability measure Λ on $\mathcal{C}(f)$ such that for all $t \in [k]$, and a uniformly random choice of S with $|S| = t$, $\pi : S \to S$ and $b \in \{-1,1\}^S$, the following signed measure on $(t+1) \times (t+1)$ matrices:*

$$\Lambda^{(t)} \;:=\; \mathop{\mathbb{E}}_{|S|=t} \; \mathop{\mathbb{E}}_{\pi:S\to S} \; \mathop{\mathbb{E}}_{b\in\{-1,1\}^{|S|}} \left[\hat{f}(S) \cdot \left(\prod_{i \in S} b_i \right) \cdot \Lambda_{S,\pi,b} \right]$$

is identically zero.

We refer to a measure Λ which satisfies the above condition, as a vanishing measure. We will obtain this measure by considering limits of the various strategies for Harry in the games $\mathcal{G}_{p,q}$. We first consider the limit for each p as $q \to \infty$.

Lemma 3.7 *For each $p \in \mathbb{N}$, there exists a limiting distribution Λ_p over $\mathcal{R}_p$ such that for every q and Γ_q,*

$$\mathsf{PayOff}(\Lambda_p, \Gamma_q) \;\le\; r_p\,,$$

where $r_p = \lim_{q\to\infty} \mathcal{V}(p,q)$.

Proof: For any row p we have a sequence of distributions $\{\Lambda_{p,q}\}_{q\in\mathbb{N}}$ such that for all q and Γ_q,

$$\mathsf{PayOff}(\Lambda_{p,q}, \Gamma_q) \;\le\; \mathcal{V}(p,q) \;\le\; r_p$$

where the second inequality used the fact that the numbers $\mathcal{V}(p,q)$ are non-decreasing in q (see Lemma 3.3). Also, for a fixed p, each $\Lambda_{p,q}$ can be viewed as a vector in $[0,1]^{|\mathcal{R}_p|}$ where $\mathcal{R}_p$ is the class of distributions over R_p with probabilities being integer multiples of $1/2^p$ (and thus

$|\mathcal{R}_p| \leq (2^p+1)^{|R_p|}$). Hence by the Bolzano-Weierstrass Theorem, the sequence $\{\Lambda_{p,q}\}_{q\in\mathbb{N}}$ has a convergent subsequence with a limit point, which we take to be Λ_p. Since each strategy Γ_q can also be viewed as a strategy $\Gamma_{q'}$ for any $q' \geq q$, we have that $\mathsf{PayOff}(\Lambda_{p,q'}, \Gamma_q) \leq r_p$. Taking the limit as $q' \to \infty$ according to the above convergent subsequence, we have that for all q and Γ_q,

$$\mathsf{PayOff}(\Lambda_p, \Gamma_q) \;=\; \lim_{q'\to\infty} \mathsf{PayOff}(\Lambda_{p,q'}, \Gamma_q) \;\leq\; r_p\,.$$

■

For the remainder of this section, we will consider the expression of the pay-off function without the absolute value. For Λ which is a *distribution over distributions* on $\mathcal{C}_\delta(f)$ and for Γ which is a distribution over functions $\psi : \mathbb{R}^d \to V_k$, we define $\mathsf{Eval}(\Lambda, \Gamma)$ as the pay-off expression without the absolute values.

$$\mathsf{Eval}(\Lambda,\Gamma) \;:=\; \mathop{\mathbb{E}}_{\lambda\sim\Lambda}\;\mathop{\mathbb{E}}_{\psi\sim\Gamma}\;\mathop{\mathbb{E}}_{\zeta\sim\lambda}\;\mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)}\left[\sum_{S\neq\emptyset}\hat{f}(S)\cdot\prod_{i\in S}\psi(\mathbf{y}_i)\right].$$

As before, for a fixed function ψ, we use $\mathsf{Eval}(\Lambda, \psi)$ to denote the expression where we omit the outer expectation over Γ. The expression $\mathsf{Eval}(\zeta, \psi)$ is defined similarly for a fixed $\zeta \in \mathcal{C}_\delta(f)$. Note that for Λ_p as given by Lemma 3.7 and any Γ_q, we have that

$$|\mathsf{Eval}(\Lambda_p,\Gamma_q)| \;\leq\; \mathop{\mathbb{E}}_{\lambda\sim\Lambda_p}\;\mathop{\mathbb{E}}_{\psi\sim\Gamma_q}\left|\mathop{\mathbb{E}}_{\zeta\sim\lambda}\;\mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)}\left[\sum_{S\neq\emptyset}\hat{f}(S)\cdot\prod_{i\in S}\psi(\mathbf{y}_i)\right]\right| \;\leq\; r_p\,.$$

Since for the purpose of computing $\mathsf{Eval}(\Lambda_p, \Gamma_q)$, we can merge the expectations over $\lambda \sim \Lambda$ and $\zeta \sim \lambda$, we will now simply consider each Λ_p to be a probability measure over $R_p \subseteq \mathcal{C}_\delta(f)$. We will obtain the desired probability measure Λ by taking a limit of the measures Λ_p obtained above. However, since the measures Λ_p are supported on sets R_p with growing size, we will need to be somewhat careful in taking the limit and will use the weak* topology to do so.

Since $\lim_{p\to\infty} r_p = L = 0$, the function $\lim_{p\to\infty} \mathsf{Eval}(\Lambda_p, \Gamma_q) = 0$ for any Γ_q. In particular, $\lim_{p\to\infty} \mathsf{Eval}(\Lambda_p, \psi_q) = 0$ for any $q \in \mathbb{N}$ and function $\psi_q : \mathbb{R}^d \to V_k$ which is constant on the cells of the partition $\mathcal{P}_q$ and is 0 outside the box $[-1, 1]^d$. We use this to prove the following lemma [11].

Lemma 3.8 *There exits a probability measure Λ on $\mathcal{C}_\delta(f)$ such that for all $q \in \mathbb{N}$ and all functions ψ_q, we have $\mathsf{Eval}(\Lambda, \psi_q) = 0$.*

Proof: Note that $\mathcal{C}_\delta(f)$ is a closed and bounded subset of $\mathbb{R}^{(k+1)^2}$ and is hence compact by the Heine-Borel Theorem. Also, by Theorem 2.9, we have that the space of probability measures on $\mathcal{C}_\delta(f)$ is compact and metrizable in the weak* topology.

From the compactness, we obtain that the infinite sequence $\{\Lambda_p\}_{p\in\mathbb{N}}$ (viewed as a sequence of probability measures on $\mathcal{C}_\delta(f)$) has a convergent subsequence with a limit point, say Λ. By the

[11] We remark that Lemma 3.8 is the main reason why we need to have an absolute value in the pay-off of our games (and hence restrict ourselves to strong approximation resistance). Without the absolute value, we can only prove $\mathsf{Eval}(\Lambda, \psi_q) \leq 0$ for all ψ_q, which does not suffice for our purpose.

definition of weak* topology, we have that for any continuous function $h : \mathcal{C}_\delta(f) \to \mathbb{R}$, taking a limit over the above subsequence, we get

$$\lim_{p\to\infty} \int h(\zeta) d\Lambda_p(\zeta) \;=\; \int h(\zeta) d\Lambda(\zeta)\,.$$

Also, note that by Lemma 3.4 the function $\mathsf{Eval}(\zeta, \psi)$ is $O_{k,d,\delta}(1)$-Lipschitz continuous, when viewed as a function of ζ. Hence, taking limits according to the above subsequence, we get that for all $q \in \mathbb{N}$ and functions ψ_q

$$0 \;=\; \lim_{p\to\infty} \mathop{\mathbb{E}}_{\zeta\sim\Lambda_p} [\mathsf{Eval}(\zeta, \psi_q)] \;=\; \mathop{\mathbb{E}}_{\zeta\sim\Lambda} [\mathsf{Eval}(\zeta, \psi_q)] \;=\; \mathsf{Eval}(\Lambda, \psi_q)$$

as claimed. ∎

To show that this implies the properties claimed in Theorem 3.6 for the limiting measure Λ, we think of the function $\mathsf{Eval}(\Lambda, \psi)$ defined as

$$\mathsf{Eval}(\Lambda, \psi) \;=\; \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} \left[\sum_{S\neq\emptyset} \hat{f}(S) \cdot \prod_{i\in S} \psi(\mathbf{y}_i) \right],$$

as a degree-k "polynomial" in the (infinite set of) variables $\psi(\mathbf{y})$ for all $\mathbf{y} \in \mathbb{R}^d$. The intuition is that since the polynomial takes the value 0 for all "assignments" ψ to the variables, all its "coefficients" must be zero.

Of course, the above is not a formal argument since the number of variables is infinite. To formalize this, we define the following quantity, which plays the role of the "coefficient" for the term $\prod_{i=1}^t \psi(\mathbf{y}_i)$ for $t \le k$. Note that in the expression for $\mathsf{Eval}(\Lambda, \psi)$, the term $\prod_{i=1}^t \psi(\mathbf{y}_i)$ can arise for any $S \subseteq [k]$ with $|S| = t$ i.e., $(\mathbf{y}_1, \ldots, \mathbf{y}_t)$ can be any ordering of any subset of size t for the points $\mathbf{z}_1, \ldots, \mathbf{z}_k$ which we sample for computing the pay-off. Also, we can also get a term involving $\psi(\mathbf{y}_i)$ if $\mathbf{y}_i = -\mathbf{z}_j$ for some $j \in [k]$, since we have the constraints $\psi(-\mathbf{z}) = -\psi(\mathbf{z})$. Taking these into account, we define the following.

Definition 3.9 *We define $\theta^{(t)} : (\mathbb{R}^d)^t \to \mathbb{R}$ on formal variables $\{\mathbf{y}_1, ..., \mathbf{y}_t\}$ as follows:*

$$\theta^{(t)}(\mathbf{y}_1, \ldots, \mathbf{y}_t) \;:=\; \sum_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \left[\hat{f}(S) \cdot \left(\prod_{i=1}^t b_i \right) \cdot \gamma_{t,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_t), \zeta_{S,\pi,b}\right) \right], \quad (3.4)$$

where $\gamma_{t,d}(\cdot, \zeta_{S,\pi,b})$ is the joint density of t correlated Gaussians in $\mathbb{R}^d$, with different coordinates being independent and the moments for each coordinate given by the appropriate submatrix ζ_S of ζ permuted according to π and modified according to the signs specified by b. Also, Λ is the limiting measure as above.

The following properties follow easily from the definition of the function $\theta^{(t)}$.

Claim 3.10 *For all $t \in [k]$ and for all $(\mathbf{y}_1, \ldots, \mathbf{y}_t) \in (\mathbb{R}^d)^t$, we have that*

- *For all permutations $\pi' : [t] \to [t]$, $\theta^{(t)}(\pi'(\mathbf{y}_1, \ldots, \mathbf{y}_t)) = \theta^{(t)}(\mathbf{y}_1, \ldots, \mathbf{y}_t)$.*
- *For all $b' \in \{-1, 1\}^t$, $\theta^{(t)}(b'_1\mathbf{y}_1, \ldots, b'_t\mathbf{y}_t) = \left(\prod_{i=1}^t b'_i\right) \cdot \theta^{(t)}(\mathbf{y}_1, \ldots, \mathbf{y}_t)$*

Proof: By definition of $\zeta_{S,\pi,b}$, we have $\gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t),\zeta_{S,\pi,b}\right) = \gamma_{t,d}\left(\pi(b_1\mathbf{y}_1,\ldots,b_t\mathbf{y}_t),\zeta_S\right)$. We can then write

$$\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t) \;=\; \sum_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \left[\hat{f}(S)\cdot\left(\prod_{i=1}^{t} b_i\right)\cdot\gamma_{t,d}\left(\pi(b_1\mathbf{y}_1,\ldots,b_t\mathbf{y}_t),\zeta_S\right)\right].$$

Since the expression already involves expectation over all permutations π of each tuple, replacing $(\mathbf{y}_1,\ldots,\mathbf{y}_t)$ by $\pi(\mathbf{y}_1,\ldots,\mathbf{y}_t)$ does not change the value of the function. Similarly, for any $b' \in \{-1,1\}^t$, we get

$$\begin{aligned}\theta^{(t)}(b'_1\mathbf{y}_1,\ldots,b'_t\mathbf{y}_t) \;&=\; \sum_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \left[\hat{f}(S)\cdot\left(\prod_{i=1}^{t} b_i\right)\cdot\gamma_{t,d}\left(\pi(b_1b'_1\mathbf{y}_1,\ldots,b_tb'_t\mathbf{y}_t),\zeta_S\right)\right]\\ &=\; \sum_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \left[\hat{f}(S)\cdot\left(\prod_{i=1}^{t} b_ib'_i\right)\cdot\gamma_{t,d}\left(\pi(b_1\mathbf{y}_1,\ldots,b_t\mathbf{y}_t),\zeta_S\right)\right],\end{aligned}$$

which equals $\left(\prod_{i=1}^{t} b'_i\right)\cdot\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)$ as claimed. ∎

The next claim shows that the functions $\theta^{(t)}$ indeed provide the right notion of "coefficients" when we think of the function $\mathsf{Eval}(\Lambda,\psi)$ as a polynomial in the values $\psi(\mathbf{z})$.

Claim 3.11 *Let Λ be the measure as above and let $\psi : \mathbb{R}^d \to V_k$ be a measurable odd function. Then,*

$$\mathsf{Eval}(\Lambda,\psi) \;=\; \sum_{t=1}^{k} \int_{(\mathbb{R}^d)^t} \theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)\cdot\left(\prod_{i=1}^{t}\psi(\mathbf{y}_i)\right) d\mathbf{y}_1\ldots d\mathbf{y}_t\,.$$

Proof: Since ψ is assumed to be measurable and the integral above is bounded, we will freely switch the order of integrals in the argument below. We have

$$\begin{aligned}&\mathsf{Eval}(\Lambda,\psi)\\ &=\; \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \mathop{\mathbb{E}}_{\mathbf{z}_1,\ldots,\mathbf{z}_k\sim\mathcal{N}_d(\zeta)} \left[\sum_{S\neq\emptyset}\hat{f}(S)\cdot\prod_{i\in S}\psi(\mathbf{z}_i)\right]\\ &=\; \sum_{t=1}^{k}\sum_{|S|=t}\hat{f}(S)\cdot \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_t\sim\mathcal{N}_d(\zeta_S)} \left[\prod_{i=1}^{t}\psi(\mathbf{y}_i)\right]\\ &=\; \sum_{t=1}^{k}\sum_{|S|=t}\hat{f}(S)\cdot \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \int_{(\mathbb{R}^d)^t} \gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t),\zeta_S\right)\cdot\left(\prod_{i=1}^{t}\psi(\mathbf{y}_i)\right) d\mathbf{y}_1\ldots d\mathbf{y}_t\\ &=\; \sum_{t=1}^{k}\sum_{|S|=t}\hat{f}(S)\cdot \mathop{\mathbb{E}}_{b\in\{-1,1\}^t}\mathop{\mathbb{E}}_{\zeta\sim\Lambda} \int_{(\mathbb{R}^d)^t} \gamma_{t,d}\left((b_1\mathbf{y}_1,\ldots,b_t\mathbf{y}_t),\zeta_S\right)\cdot\left(\prod_{i=1}^{t}\psi(b_i\mathbf{y}_i)\right) d\mathbf{y}_1\ldots d\mathbf{y}_t\\ &=\; \sum_{t=1}^{k}\sum_{|S|=t}\hat{f}(S)\cdot \mathop{\mathbb{E}}_{b\in\{-1,1\}^t}\mathop{\mathbb{E}}_{\zeta\sim\Lambda} \int_{(\mathbb{R}^d)^t} \gamma_{t,d}\left((b_1\mathbf{y}_1,\ldots,b_t\mathbf{y}_t),\zeta_S\right)\cdot\left(\prod_{i=1}^{t} b_i\right)\cdot\left(\prod_{i=1}^{t}\psi(\mathbf{y}_i)\right) d\mathbf{y}_1\ldots d\mathbf{y}_t\,,\end{aligned}$$

where the last equality used the fact that ψ is odd. Finally, we note that the term $\prod_{i=1}^{t}\psi(\mathbf{y}_i)$ can arise from any permutation of the tuple $(\mathbf{y}_1,\ldots,\mathbf{y}_t)$. We thus re-write the expression above as

$$\begin{aligned}
&\mathsf{Eval}(\Lambda,\psi)\\
&=\sum_{t=1}^{k}\int_{(\mathbb{R}^d)^t}\sum_{|S|=t}\mathop{\mathbb{E}}_{\substack{\pi:[t]\to[t]\\ b\in\{-1,1\}^t}}\mathop{\mathbb{E}}_{\zeta\sim\Lambda}\left[\hat{f}(S)\cdot\left(\prod_{i=1}^{t}b_i\right)\cdot\gamma_{t,d}\left(\pi(b_1\mathbf{y}_1,\ldots,b_t\mathbf{y}_t),\zeta_S\right)\right]\cdot\left(\prod_{i=1}^{t}\psi(\mathbf{y}_i)\right)d\mathbf{y}_1\ldots d\mathbf{y}_t\\
&=\sum_{t=1}^{k}\int_{(\mathbb{R}^d)^t}\sum_{|S|=t}\mathop{\mathbb{E}}_{\substack{\pi:S\to S\\ b\in\{-1,1\}^S}}\mathop{\mathbb{E}}_{\zeta\sim\Lambda}\left[\hat{f}(S)\cdot\left(\prod_{i\in S}b_i\right)\cdot\gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t),\zeta_{S,\pi,b}\right)\right]\cdot\left(\prod_{i=1}^{t}\psi(\mathbf{y}_i)\right)d\mathbf{y}_1\ldots d\mathbf{y}_t\\
&=\sum_{t=1}^{k}\int_{(\mathbb{R}^d)^t}\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)\cdot\left(\prod_{i=1}^{t}\psi(\mathbf{y}_i)\right)d\mathbf{y}_1\ldots d\mathbf{y}_t\,,
\end{aligned}$$

as claimed. ■

We next show that $\theta^{(t)}$ is a "nice" function. For this we shall need to use the fact that Λ is a measure over $\mathcal{C}_\delta(f)$, and that matrices in $\mathcal{C}_\delta(f)$ have each eigenvalue at least δ.

Lemma 3.12 *For all $t\in[k]$, $\theta^{(t)}$ is bounded i.e., $\left\|\theta^{(t)}\right\|_\infty\le O_{k,d,\delta}(1)$, and it is $O_{k,d,\delta}(1)$-Lipschitz.*

Proof: We first argue that $\theta^{(t)}$ is bounded. The Gaussian density $\gamma_{t,d}\left(\cdot,\zeta_{S,\pi,b}\right)$ is at most $\frac{1}{(2\pi)^{td/2}|\Sigma|^{d/2}}$ where Σ is the covariance matrix associated with $\zeta_{S,\pi,b}$ with $\Sigma_{ij}=\zeta_{S,\pi,b}(i,j)-\zeta_{S,\pi,b}(0,i)\cdot\zeta_{S,\pi,b}(0,j)$, and $|\Sigma|$ denotes the determinant. Since $\zeta\in\mathcal{C}_\delta(f)$, all the eigenvalues of Σ are at least δ and hence $|\Sigma|\ge\delta^t$. Also, since $\left|\hat{f}(S)\right|\le 1$, we get

$$\left\|\theta^{(t)}\right\| \;\le\; \binom{k}{t}\cdot\frac{1}{(2\pi)^{td/2}\cdot\delta^{td/2}} \;\le\; \frac{1}{\delta^{kd/2}}\,.$$

Let Σ be the covariance matrix as above and μ be the vector of means with $\mu_i=\zeta_{S,\pi,b}(0,i)$. Also, for $l\in[d]$, let $\mathbf{y}^{(l)}\in\mathbb{R}^t$ denote the vector $((\mathbf{y}_1)_l,\ldots,(\mathbf{y}_t)_l)$ obtained by taking the l^{th} coordinates of $\mathbf{y}_1,\ldots,\mathbf{y}_t$. The Gaussian density $\gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t),\zeta_{S,\pi,b}\right)$ can then be written as

$$\gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t),\zeta_{S,\pi,b}\right) \;=\; \prod_{l=1}^{d}\left(\frac{1}{(2\pi)^{t/2}\,|\Sigma|^{1/2}}\cdot\exp\left(-\frac{1}{2}\cdot(\mathbf{y}^{(l)}-\mu)^T\Sigma^{-1}(\mathbf{y}^{(l)}-\mu)\right)\right).$$

The density is a function on $\mathbb{R}^{dt}$. The gradient on the coordinates corresponding to $\mathbf{y}^{(l)}$ can be written as

$$\left[\nabla\left(\gamma_{t,d}\left(\mathbf{y},\zeta_{S,\pi,b}\right)\right)\right]_l \;=\; \frac{1}{(2\pi)^{td/2}\cdot|\Sigma|^{d/2}}\cdot\left(-\Sigma^{-1}(\mathbf{y}^{(l)}-\mu)\right)\cdot\prod_{l=1}^{d}\exp\left(-\frac{1}{2}\cdot(\mathbf{y}^{(l)}-\mu)^T\Sigma^{-1}(\mathbf{y}^{(l)}-\mu)\right).$$

Since Σ^{-1} is positive semidefinite, we can define a matrix $\Sigma^{-1/2}$. Also, since Σ has eigenvalues at least δ, we have $\left\|\Sigma^{-1}(\mathbf{y}^{(l)}-\mu)\right\|\le\frac{1}{\sqrt{\delta}}\cdot\left\|\Sigma^{-1/2}(\mathbf{y}^{(l)}-\mu)\right\|$. Using this, we can bound the norm of gradient in the coordinates corresponding to $\mathbf{y}^{(l)}$ as

$$\begin{aligned}
\left\|\left[\nabla\left(\gamma_{t,d}\left(\mathbf{y},\zeta_{S,\pi,b}\right)\right)\right]_l\right\| &= \frac{1}{(2\pi)^{td/2}\cdot|\Sigma|^{d/2}}\cdot\left\|\Sigma^{-1}(\mathbf{y}^{(l)}-\mu)\right\|\cdot\prod_{l=1}^{d}\exp\left(-\frac{1}{2}\cdot\left\|\Sigma^{-1/2}(\mathbf{y}^{(l)}-\mu)\right\|^2\right)\\
&\le \frac{\left\|\Sigma^{-1/2}(\mathbf{y}^{(l)}-\mu)\right\|}{(2\pi)^{td/2}\cdot|\Sigma|^{d/2}\cdot\sqrt{\delta}}\cdot\exp\left(-\frac{1}{2}\cdot\sum_{l=1}^{d}\left\|\Sigma^{-1/2}(\mathbf{y}^{(l)}-\mu)\right\|^2\right).
\end{aligned}$$

This bounds the norm of the gradient as

$$\begin{aligned}\|\nabla\left(\gamma_{t,d}\left(y,\zeta_{S,\pi,b}\right)\right)\|^2 &\leq \frac{\sum_{l=1}^{d}\left\|\Sigma^{-1/2}(\mathbf{y}^{(l)}-\mu)\right\|^2}{(2\pi)^{td}\cdot|\Sigma|^d\cdot\delta}\cdot\exp\left(-\sum_{l=1}^{d}\left\|\Sigma^{-1/2}(\mathbf{y}^{(l)}-\mu)\right\|^2\right)\\ &\leq \frac{1}{(2\pi)^{td}\cdot|\Sigma|^d\cdot\delta}\,,\end{aligned}$$

where we used the fact that the function $x\cdot\exp(-x)$ is bounded above by 1. Using the above, we obtain a bound on the gradient of $\theta^{(t)}$ as

$$\left\|\nabla\theta^{(t)}\right\| \leq \binom{k}{t}\cdot\frac{1}{(2\pi)^{td/2}\cdot\delta^{(td+1)/2}} \leq \frac{1}{\delta^{(kd+1)/2}}\,.$$

Hence, $\theta^{(t)}$ is C-Lipschitz, with $C\leq(1/\delta)^{(kd+1)/2}$. ■

Using the above properties and the fact that $\mathsf{Eval}(\Lambda,\psi_q)=0$ for all $q\in\mathbb{N}$ and all functions ψ_q, we can in fact show that the functions $\theta^{(t)}$ must in fact be identically zero on the entire box $([-1,1]^d)^t$.

Lemma 3.13 *For all $t\in[k]$ and all $\mathbf{y}_1,\ldots,\mathbf{y}_t\in[-1,1]^d$, we have $\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)=0$.*

Proof: Let H denote the space $[0,1]\times[-1,1]^{d-1}$. By Claim 3.10, we only need to show $\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)=0$ for all $\mathbf{y}_1,\ldots,\mathbf{y}_t\in H$, since changing the sign of any input $\mathbf{y}_i$ only changes the sign of $\theta^{(t)}$. Also, by Claim 3.11, we have that for any odd function $\psi_q:\mathbb{R}^d\to V_k$, which is 0 outside $[-1,1]^d$,

$$\begin{aligned}\mathsf{Eval}(\Lambda,\psi_q) &= \sum_{t=1}^{k}\int_{(\mathbb{R}^d)^t}\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)\cdot\left(\prod_{i=1}^{t}\psi_q(\mathbf{y}_i)\right)d\mathbf{y}_1\ldots d\mathbf{y}_t\\ &= \sum_{t=1}^{k}\int_{([-1,1]^d)^t}\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)\cdot\left(\prod_{i=1}^{t}\psi_q(\mathbf{y}_i)\right)d\mathbf{y}_1\ldots d\mathbf{y}_t\\ &= \sum_{t=1}^{k}2^t\cdot\int_{H^t}\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)\cdot\left(\prod_{i=1}^{t}\psi_q(\mathbf{y}_i)\right)d\mathbf{y}_1\ldots d\mathbf{y}_t\,.\end{aligned}$$

The second equality above used the fact that ψ_q is 0 outside $[-1,1]^d$. The last equality used that by Claim 3.10 and the fact that ψ_q is odd, we have for any $b\in\{-1,1\}^t$

$$\theta^{(t)}(b_1\mathbf{y}_1,\ldots,b_t\mathbf{y}_t)\cdot\left(\prod_{i=1}^{t}\psi_q(b_i\mathbf{y}_i)\right) = \theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)\cdot\left(\prod_{i=1}^{t}\psi_q(\mathbf{y}_i)\right)\,.$$

Recall that for each $q\in\mathbb{N}$ the functions ψ_q are constant on the cells of the partition $\mathcal{P}_q$ which divides $[-1,1]^d$ in $2^{(q+1)d}$ equal-sized boxes. By the above expression for $\mathsf{Eval}(\Lambda,\psi_q)$ and Lemma 3.8, we have that for any such function ψ_q

$$\sum_{t=1}^{k}2^t\cdot\int_{H^t}\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t)\cdot\left(\prod_{i=1}^{t}\psi_q(\mathbf{y}_i)\right)d\mathbf{y}_1\ldots d\mathbf{y}_t = 0\,.$$

The partition $\mathcal{P}_q$ induces a partition $\mathcal{P}^{(t)}$ on H^t such that $\prod_{i=1}^t \psi_q(\mathbf{y}_i)$ is constant on each cell of the partition $\mathcal{P}^{(t)}$. We will use $w \in \mathcal{P}^{(t)}$ to denote a cell of this partition. Also, note that the cell w can be written as $(w_1, \ldots, w_t)$, where each w_i denotes a cell in $\mathcal{P}_q$.

We define the function $\overline{\theta}^{(t)}$, which is $\theta^{(t)}$ averaged over each cell of $\mathcal{P}^{(t)}$ (which has volume 2^{-qdt})

$$\overline{\theta}^{(t)}(w) \;:=\; (2^{qd})^t \cdot \int_{\mathbf{y}' \in w} \theta^{(t)}(\mathbf{y}'_1, \ldots, \mathbf{y}'_t) d\mathbf{y}'_1 \ldots d\mathbf{y}'_t \,.$$

Also, since $\prod_{i=1}^t \psi_q(\mathbf{y}_i)$ is constant on each w, we will use $\prod_{i=1}^t \psi_q(w_i)$ to denote its value over the cell w. Using the above, we get

$$\mathsf{Eval}(\Lambda, \psi_q) \;=\; \sum_{t=1}^{k} 2^t \cdot \sum_{w \in \mathcal{P}^{(t)}} 2^{-qdt} \cdot \overline{\theta}^{(t)}(w_1, \ldots, w_t) \cdot \left(\prod_{i=1}^{t} \psi_q(w_i) \right) \;=\; 0 \,,$$

for all functions ψ_q. Since each ψ_q is defined by $2^{(q+1)d}/2$ values, corresponding to the cells of $\mathcal{P}_q$ in H, the above can be viewed as a degree-k polynomial in $2^{(q+1)d}/2$ variables. Note that $\prod_{i=1}^t \psi_q(w_i)$ can arise from any permutation of the tuple $(w_1, \ldots, w_t)$. Since $\theta^{(t)}$ is invariant under the permutation of its inputs by Claim 3.10, the coefficient of $\prod_{i=1}^t \psi_q(w_i)$ is

$$2^t \cdot 2^{-qdt} \cdot C_{t,w} \cdot \overline{\theta}^{(t)}(w_1, \ldots, w_t) \,,$$

where $C_{t,w}$ is the number of permutations of the tuple $(w_1, \ldots, w_t)$.

We have that the above polynomial over $\mathbb{R}$ is zero for all assignments to its variables from the set $V_k = \{\frac{r}{k} \;:\; -k \le r \le k\}$. However, from the Schwartz-Zippel lemma, we know that a non-zero degree-k polynomial must only take the value 0 with probability $\frac{k}{|V_k|} < \frac{1}{2}$, over a random assignment to its variables from the set V_k. This means that the polynomial above must be identically zero and hence

$$\forall t \in [k], \quad \forall (w_1, \ldots, w_t) \in \mathcal{P}^{(t)} \qquad \overline{\theta}^{(t)}(w_1, \ldots, w_t) \;=\; 0 \,.$$

Each cell of the partition $\mathcal{P}^{(t)}$ is a box in $\mathbb{R}^{dt}$ with each side having length 2^{-q}. Since $\overline{\theta}^{(t)}(w_1, \ldots, w_t)$ is the average of $\theta^{(t)}$ over the box corresponding to $(w_1, \ldots, w_t)$ and $\theta^{(t)}$ is $O_{k,d,\delta}(1)$-Lipschitz by Lemma 3.12, we have that for some constant $C_{k,d,\delta}$

$$\forall t \in [k], \quad \forall (\mathbf{y}_1, \ldots, \mathbf{y}_t) \in H^t \qquad \left| \theta^{(t)}(\mathbf{y}_1, \ldots, \mathbf{y}_t) \right| \;\le\; \frac{C_{k,d,\delta}}{2^q} \,.$$

However, since the above holds for all $q \in \mathbb{N}$, we must have that $\theta^{(t)}(\mathbf{y}_1, \ldots, \mathbf{y}_t) = 0$ for all $(\mathbf{y}_1, \ldots, \mathbf{y}_t) \in H^t$ and hence for all $(\mathbf{y}_1, \ldots, \mathbf{y}_t) \in ([-1, 1]^d)^t$. ∎

For a set S with $|S| = t$, permutation $\pi : [t] \to [t]$ and $b \in \{-1, 1\}^t$, let $\Lambda_{S,\pi,b}$ denote the projection of Λ to $(t+1) \times (t+1)$ matrices as defined in Section 2. We define the following signed measure on space of $(t+1) \times (t+1)$ matrices

$$\Lambda^{(t)} \;:=\; \mathop{\mathbb{E}}_{|S|=t} \; \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \; \mathop{\mathbb{E}}_{b \in \{-1,1\}^t} \left[\hat{f}(S) \cdot \left(\prod_{i=1}^{t} b_i \right) \cdot \Lambda_{S,\pi,b} \right]$$

Lemma 3.13 immediately gives the following. Note that the integration below is over $\zeta' \sim \Lambda^{(t)}$ and the tuple $\mathbf{y}_1, \ldots, \mathbf{y}_t$ is fixed.

Claim 3.14 *For all $t \in [k]$ and for all $\mathbf{y}_1, \ldots, \mathbf{y}_t \in [-1,1]^d$, we have*

$$\int \gamma_{t,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_t), \zeta'\right) d\Lambda^{(t)}(\zeta') \;=\; 0\,.$$

Proof: We start by expanding the expression for $\theta^{(t)}$.

$$\begin{aligned}
\theta^{(t)}(\mathbf{y}_1, \ldots, \mathbf{y}_t) &= \binom{k}{t} \cdot \mathop{\mathbb{E}}_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \int \hat{f}(S) \cdot \left(\prod_{i=1}^{t} b_i\right) \cdot \gamma_{t,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_t), \zeta_{S,\pi,b}\right) d\Lambda(\zeta) \\
&= \binom{k}{t} \cdot \mathop{\mathbb{E}}_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \int \hat{f}(S) \cdot \left(\prod_{i=1}^{t} b_i\right) \cdot \gamma_{t,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_t), \zeta'\right) d\Lambda_{S,\pi,b}(\zeta') \\
&= \binom{k}{t} \cdot \int \gamma_{t,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_t), \zeta'\right) d\Lambda^{(t)}(\zeta')\,.
\end{aligned}$$

The claim follows by using that $\theta^{(t)}(\mathbf{y}_1, \ldots, \mathbf{y}_t) = 0$ for all $\mathbf{y}_1, \ldots, \mathbf{y}_t \in [-1,1]^d$ by Lemma 3.13. ■

From the claim we get that the integral of $\gamma_{t,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_t), \zeta'\right)$ with respect to the signed measure $\Lambda^{(t)}$ is zero for all $\mathbf{y}_1, \ldots, \mathbf{y}_t \in [-1,1]^d$. We will use it to show that the integral of all continuous functions must be zero with respect to $\Lambda^{(t)}$ and hence $\Lambda^{(t)}$ must itself be identically zero. However, we will need to modify $\Lambda^{(t)}$ a little to prove this.

We begin by considering the expression for $\gamma_{t,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_t), \zeta'\right)$. We note that there is a bijection between the matrices ζ' and the pairs (Σ, μ), where $\mu \in \mathbb{R}^t$ is a vector of means with $\mu_i = \zeta'(0,i)$ and Σ is the $t \times t$ covariance matrix with $\Sigma_{ij} = \zeta'(i,j) - \mu_i \cdot \mu_j$. Also, since $\zeta' = \zeta_{S,\pi,b}$ for some $\zeta \in \mathcal{C}_\delta(f)$, we have that Σ is an invertible matrix with each eigenvalue at least δ. We shall use M to denote the matrix Σ^{-1} which has all eigenvalues at most $1/\delta$. Also, as before, for vectors $\mathbf{y}_1, \ldots, \mathbf{y}_t \in \mathbb{R}^d$, and for $l \in [d]$, we use $\mathbf{y}^{(l)} \in \mathbb{R}^t$ to denote the vector consisting of the l^{th} coordinates of $\mathbf{y}_1, \ldots, \mathbf{y}_t$. We can then write

$$\begin{aligned}
&\gamma_{t,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_t), \zeta'\right) \\
&= \frac{1}{(2\pi)^{td/2} \cdot |\Sigma|^{d/2}} \cdot \exp\left(-\frac{1}{2} \cdot \sum_{l=1}^{d} (\mathbf{y}^{(l)} - \mu)^T M (\mathbf{y}^{(l)} - \mu)\right) \\
&= \frac{1}{(2\pi)^{td/2} \cdot |\Sigma|^{d/2}} \cdot \exp\left(-\frac{1}{2} \sum_{i,j=1}^{t} M_{ij} \langle \mathbf{y}_i, \mathbf{y}_j \rangle - \frac{d}{2} \sum_{i,j=1}^{t} M_{ij}\mu_i\mu_j + \sum_{i,j=1}^{t} M_{ij}\mu_j \langle \mathbf{y}_i, \mathbf{1} \rangle\right) \\
&= \gamma_{t,d}\left((\mathbf{0}, \ldots, \mathbf{0}), \zeta'\right) \cdot \exp\left(-\frac{1}{2} \sum_{i,j=1}^{t} M_{ij} \langle \mathbf{y}_i, \mathbf{y}_j \rangle + \sum_{i,j=1}^{t} M_{ij}\mu_j \langle \mathbf{y}_i, \mathbf{1} \rangle\right)\,,
\end{aligned}$$

where $\mathbf{1} \in \mathbb{R}^d$ denotes the vector $(1, \ldots, 1)$ and $\mathbf{0} \in \mathbb{R}^d$ denotes the vector $(0, \ldots, 0)$.

We will try to argue that for $d \geq k+1$, the values $\{\langle \mathbf{y}_i, \mathbf{y}_j \rangle\}_{i,j\in[t]}$ and $\{\langle \mathbf{y}_i, \mathbf{1} \rangle\}_{i\in[t]}$ are "independent enough" so that if the integral of $\gamma_{t,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_t), \zeta'\right)$ with respect to $\Lambda^{(t)}$ vanishes for all $\mathbf{y}_1, \ldots, \mathbf{y}_t \in [-1,1]^d$, then $\Lambda^{(t)}$ vanishes. However, the values $\{\langle \mathbf{y}_i, \mathbf{y}_j \rangle\}_{i,j\in[t]}$ and $\{\langle \mathbf{y}_i, \mathbf{1} \rangle\}_{i\in[t]}$ cannot vary completely independently, since they are required to form a positive semidefinite matrix. To handle this, we define the variables (for $\beta > 0$ to be chosen later)

$$X_{ij} = \begin{cases} \langle \mathbf{y}_i, \mathbf{y}_j \rangle & \text{if } i \neq j \\ \langle \mathbf{y}_i, \mathbf{y}_i \rangle - \beta & \text{if } i = j \end{cases} \quad \text{and} \quad Z_i = \langle \mathbf{y}_i, \mathbf{1} \rangle\,. \tag{3.5}$$

Let N denote the vector $\Sigma^{-1}\mu = M\mu$. We can then write

$$\gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t),\zeta'\right) \;=\; \gamma_{t,d}\left((\mathbf{0},\ldots,\mathbf{0}),\zeta'\right)\cdot\exp\left(-\frac{\beta}{2}\cdot Tr(M)\right)\cdot\exp\left(-\frac{1}{2}(M\bullet X)+\langle N,Z\rangle\right),$$

where $M \bullet X$ denotes the Frobenius inner product of the two matrices.

Note that there is a bijection between the pairs (M,N) and the pairs (Σ,μ), and hence also between the pairs (M,N) and the matrices ζ'. We can then view the expression $\gamma_{t,d}\left((\mathbf{0},\ldots,\mathbf{0}),\zeta'\right)\cdot\exp\left(-\frac{\beta}{2}\cdot Tr(M)\right)$ as a function of the pair (M,N), say $g_\beta^{(t)}(M,N)$. Also viewing the Gaussian density as a function of the pair (M,N), we can write

$$\gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t),(M,N)\right) \;=\; g_\beta^{(t)}(M,N)\cdot\exp\left(-\frac{1}{2}(M\bullet X)+\langle N,Z\rangle\right).$$

Finally, note that the bijection from the pairs (Σ,μ) to the pairs (M,N) is a *continuous map*, since both the maps $\Sigma\mapsto\Sigma^{-1}$ and $\mu\mapsto\Sigma^{-1}\mu$ are continuous on the space of matrices Σ with each eigenvalue at least δ. Also, the bijection from matrices ζ' to the pairs (Σ,μ) is continuous. Thus, the bijection from matrices ζ' to the pairs (M,N) is continuous and hence maps measurable sets to measurable sets. Hence, we can also view the signed measure $\Lambda^{(t)}$ as a signed measure on the pairs (M,N).

We say that a pair (X,Z) for $X\in\mathbb{R}^{t\times t}$ and $Z\in\mathbb{R}^t$ is (β,d)-realizable if there exist $\mathbf{y}_1,\ldots,\mathbf{y}_t\in[-1,1]^d$ such that the values X_{ij} and Z_i satisfy the relation in Equation 3.5. From the above discussion and Claim 3.14, we have that for all (β,d)-realizable pairs (X,Z)

$$\int g_\beta^{(t)}(M,N)\cdot\exp\left(-\frac{1}{2}(M\bullet X)+\langle N,Z\rangle\right)d\Lambda^{(t)}(M,N) \;=\; 0\,.$$

Note that $g_\beta^{(t)}(M,N)$ is a positive valued function of the pair (M,N). Using this we define the signed measure $\tilde{\Lambda}^{(t)}$ as

$$\tilde{\Lambda}^{(t)} \;:=\; \Lambda^{(t)}\cdot g_\beta^{(t)}\,.$$

Formally, for every set A (of pairs (M,N)) in the underlying σ-algebra, we define

$$\tilde{\Lambda}^{(t)}(A) \;:=\; \int \mathbb{1}_{\{A\}}(M,N)\cdot g_\beta^{(t)}(M,N)\;d\Lambda^{(t)}(M,N)$$

This operation indeed defines a new signed measure if $g_\beta^{(t)}$ is a continuous non-negative function (see Exercise 7 in Chapter 3 of [37] for example). The required conditions on $g_\beta^{(t)}$ are easily proved.

Claim 3.15 *The function $g_\beta^{(t)}$ is a positive and continuous function of the pairs (M,N), and is bounded above by a constant $C_{k,d,\delta}$.*

Proof: Let $\zeta'(M,N)$ denote the moment matrix corresponding to (M,N). Recall that the function $g_\beta^{(t)}$ was defined as

$$g_\beta^{(t)}(M,N) \;=\; \gamma_{t,d}\left((\mathbf{0},\ldots,\mathbf{0}),\zeta'(M,N)\right)\cdot\exp\left\{-\frac{\beta}{2}\cdot Tr(M)\right\}.$$

Note that the Gaussian density $\gamma_{t,d}\left((\mathbf{0},\ldots,\mathbf{0}),\zeta'(M,N)\right)$ is a continuous function of the matrix ζ' and hence also of the pair (M,N). Also, it is positive and bounded above by $\frac{1}{(2\pi)^{td/2}}\cdot|M|^{d/2} \leq (1/\delta)^{td/2}$. Also, $\exp\left\{-\frac{\beta}{2}\cdot Tr(M)\right\}$ is a positive and continuous function of M and is bounded above by 1. Hence, their product $g_\beta^{(t)}$ is also positive, continuous and bounded as claimed. ■

From the definition of $\tilde{\Lambda}^{(t)}$, we have that for all (β,d)-realizable pairs (X,Z)

$$\int \exp\left(-\frac{1}{2}(M\bullet X)+\langle N,Z\rangle\right)d\tilde{\Lambda}^{(t)}(M,N) \;=\; 0\,.$$

The following claim shows that the class of (β,d)-realizable pairs is sufficiently rich.

Claim 3.16 *Let $X\in\mathbb{R}^{t\times t}$ be a matrix and $Z\in\mathbb{R}^t$ be a vector such that*

$$\forall i,j\in[t]\quad |X_{ij}|\leq\frac{\beta}{t+1}\qquad \textit{and}\qquad \forall i\in[t]\quad |Z_i|\leq\frac{\beta}{t+1}\,.$$

Then the pair (X,Z) is (β,d)-realizable for $d\geq k+1$ and $\beta\leq 1/2$.

Proof: Consider the $(t+1)\times(t+1)$ matrix Y defined as $Y_{00}=d$, $Y_{0i}=Y_{i0}=Z_i$ and $Y_{ii}=X_{ii}+\beta$ for $i\geq 1$, and $Y_{ij}=X_{ij}$ for $i\neq j$ when $i,j\geq 1$. The matrix is diagonally dominant and is hence positive semidefinite, when X and Z are as above.

Thus, there exist vectors $\mathbf{y}_0,\mathbf{y}_1,\ldots,\mathbf{y}_t\in\mathbb{R}^d$ when $d\geq t+1$, such that $Y_{ij}=\langle\mathbf{y}_i,\mathbf{y}_j\rangle$. Also, we have $\|\mathbf{y}_0\|^2=Y_{00}=d$ and hence we can assume (by applying a rotation if necessary) that $\mathbf{y}_0=\mathbf{1}$. Finally, we also have

$$\|\mathbf{y}_i\|^2 \;=\; Y_{ii} \;=\; X_{ii}+\beta \;\leq\; \frac{\beta}{t+1}+\beta \;\leq\; 1$$

when $\beta\leq 1/2$. Thus, all the vectors $\mathbf{y}_i$ have $\|\mathbf{y}_i\|\leq 1$ and lie in $[-1,1]^d$. By definition of the matrix Y, the pair (X,Z) satisfies the relation in Equation 3.5 and is hence (β,d)-realizable. ■

Hence, all the variables X_{ij} and Z_i are allowed to vary in a radius of $\beta/(t+1)$ and the above integral is zero for all values of these variables. We can expand the integral as a power series in these variables, and then argue that all its coefficients must be zero within the radius of convergence. However, it will be convenient to re-write the function $\exp\left\{-\frac{1}{2}(M\bullet X)+\langle N,Z\rangle\right\}$ slightly differently, before expanding it as a power series.

Note that since the matrices M and X are symmetric, the variable X_{ij} actually appears twice in X when $i\neq j$, and thus it's coefficient in $-\frac{1}{2}(M\bullet X)$ is $-M_{ij}$ when $i\neq j$ and $-M_{ii}/2$ when $i=j$. We re-write the $K=\binom{t}{2}+2t$ variables corresponding to (X,Z) as the vector $\mathbf{w}=(w_1,\ldots,w_K)$ and their coefficients as $\mathbf{a}=(a_1,\ldots,a_K)$. As before, the map from (M,N) is a continuous bijection and thus, we can view $\tilde{\Lambda}^{(t)}$ as a measure on the coefficient vectors $\mathbf{a}$. From Claim 3.16, we have

$$\int\exp(\langle\mathbf{a},\mathbf{w}\rangle)\;d\tilde{\Lambda}^{(t)}(\mathbf{a})=0\qquad\forall\mathbf{w}\in\left[-\frac{\beta}{t+1},\frac{\beta}{t+1}\right]^K\,. \tag{3.6}$$

The following bound on the coefficients will be useful.

Claim 3.17 *Let $\mathbf{a}=(a_1,\ldots,a_K)$ be as above. Then $|a_i|\leq\frac{t}{\delta}$ for each $i\in[K]$.*

Proof: The coefficients for the variables X_{ij} are $-M_{ij} = -\Sigma^{-1}_{ij}$ when $i \neq j$. Let $\mathbf{e}_i$ denote the i^{th} unit vector in the standard basis for $\mathbb{R}^t$. Then

$$\left|\Sigma^{-1}_{ij}\right| \;=\; \langle \mathbf{e}_i, \Sigma^{-1}\mathbf{e}_j\rangle \;\leq\; \left\|\Sigma^{-1}\mathbf{e}_j\right\| \;\leq\; \frac{1}{\delta}\,.$$

Similarly, the coefficient for X_{ii}, which equals $-\Sigma^{-1}_{ii}/2$ is bounded in absolute value by $\frac{1}{2\delta}$. Finally, the coefficient for Z_i equals $N_i = \left(\Sigma^{-1}\mu\right)_i$ and is bounded as

$$\left|\left(\Sigma^{-1}\mu\right)_i\right| \;\leq\; \left\|\Sigma^{-1}\mu\right\| \;\leq\; \frac{1}{\delta}\cdot\|\mu\| \;\leq\; \frac{t}{\delta}\,,$$

where the bound on $\|\mu\|$ uses that its each coordinate μ_i is in $[-1,1]$. ■

We shall expand the function $\exp(\langle \mathbf{a}, \mathbf{w}\rangle)$ as a power series and integrate each term separately to obtain a formal series $\mathcal{S}(\mathbf{w})$. To write the series, it will be convenient to use the multi-index notation. Let $\mathbf{r} = (r_1, \ldots, r_K) \in (\mathbb{Z}_+)^K$ denote a multi-index. Let $\mathbf{a}^{\mathbf{r}}$ denote the term $\prod_{i=1}^K a_i^{r_i}$ and define $\mathbf{w}^{\mathbf{r}}$ similarly. Let $|\mathbf{r}|$ denote $\sum_{i=1}^K r_i$ and let $(\mathbf{r})!$ denote $\prod_{i=1}^K (r_i!)$. Then we can write

$$\exp(\langle \mathbf{a}, \mathbf{w}\rangle) \;=\; \sum_{r=0}^{\infty} \frac{(\langle \mathbf{a}, \mathbf{w}\rangle)^r}{r!} \;=\; \sum_{\mathbf{r}\in\mathbb{Z}_+^K} \frac{\mathbf{w}^{\mathbf{r}}\cdot\mathbf{a}^{\mathbf{r}}}{(\mathbf{r})!}\,.$$

We define the series

$$\mathcal{S}(\mathbf{w}) \;=\; \sum_{\mathbf{r}\in\mathbb{Z}_+^K} \frac{\mathbf{w}^{\mathbf{r}}}{(\mathbf{r})!}\cdot\int \mathbf{a}^{\mathbf{r}}\; d\tilde{\Lambda}^{(t)}(\mathbf{a})\,.$$

We next show that this formal series converges everywhere. Using the convergence, we can equate it to the integral in Equation 3.6. The fact that the integral is zero in a box around the origin will then yield the desired conclusion.

Claim 3.18 *Let the vectors $\mathbf{a} = (a_1, \ldots, a_K)$ and the measure $\tilde{\Lambda}^{(t)}$ be as above. Then the series*

$$\mathcal{S}(\mathbf{w}) \;=\; \sum_{\mathbf{r}\in\mathbb{Z}_+^K} \frac{\mathbf{w}^{\mathbf{r}}}{(\mathbf{r})!}\cdot\int \mathbf{a}^{\mathbf{r}}\; d\tilde{\Lambda}^{(t)}(\mathbf{a})$$

is absolutely convergent for all $\mathbf{w} \in \mathbb{R}^K$.

Proof: We bound the absolute value of the integral $\int \mathbf{a}^{\mathbf{r}}\; d\tilde{\Lambda}^{(t)}(\mathbf{a})$ for each $\mathbf{r} \in \mathbb{Z}_+^K$. By claim 3.17, $|a_i| \leq \frac{t}{\delta}$ for each coordinate a_i of $\mathbf{a}$. We then have

$$\left|\int \mathbf{a}^{\mathbf{r}}\; d\tilde{\Lambda}^{(t)}(\mathbf{a})\right| \;=\; \left|\int \mathbf{a}^{\mathbf{r}}\cdot g_\beta^{(t)}(\mathbf{a})\; d\Lambda^{(t)}(\mathbf{a})\right| \;\leq\; \sup_{\mathbf{a}}\left(|\mathbf{a}^{\mathbf{r}}|\cdot g_\beta^{(t)}(\mathbf{a})\right)\cdot\int d|\Lambda^{(t)}|(\mathbf{a})\,.$$

Note that here we have used notation $|\Lambda^{(t)}|$, which is used to refer to a *positive* measure corresponding to $\Lambda^{(t)}$, which is given by the Hahn decomposition theorem for signed measures. By the decomposition theorem, any signed measure ν can be written as $\nu_+ - \nu_-$, where ν_+ and ν_- are positive measures supported on disjoint measurable sets, say P and N respectively. Then $|\nu|$ is used to refer to the measure $(\nu_+ + \nu_-)$. The inequality above follows immediately by considering this decomposition.

Also, if Λ_0 is a finite linear combination of positive measures i.e., $\Lambda_0 = \sum_i c_i \Lambda_i$, then using the above decomposition, we can say that $|\Lambda_0| \leq \sum_i |c_i| \, \Lambda_i$. By the definition of $\Lambda^{(t)}$ as a linear combination of positive measures, we can now bound the integral as

$$\int d|\Lambda^{(t)}|(\mathbf{a}) \;\leq\; \mathop{\mathbb{E}}_{|S|=t} \; \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \; \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \left[\left| \hat{f}(S) \right| \cdot \left| \prod_{i=1}^{t} b_i \right| \cdot \int d\Lambda_{S,\pi,b}(\mathbf{a}) \right] \;\leq\; 1 \,,$$

since each $\Lambda_{S,\pi,b}$ is a probability measure. Using the bound on the coefficients a_i, we have that $|\mathbf{a}^{\mathbf{r}}| \leq (t/\delta)^{|\mathbf{r}|}$. Also, by Claim 3.15, we have that $g_\beta^{(t)} \leq C_{k,d,\delta}$. Thus, we get

$$\left| \int \mathbf{a}^{\mathbf{r}} \; d\tilde{\Lambda}^{(t)}(\mathbf{a}) \right| \;\leq\; C_{k,d,\delta} \cdot \left(\frac{t}{\delta} \right)^{\mathbf{r}} .$$

For $\mathbf{w} = (w_1, \ldots, w_K) \in \mathbb{R}^K$, let $\mathbf{w}_+ = (|w_1|, \ldots, |w_K|)$ be the vector of absolute values of all the entries of $\mathbf{w}$. To show that $\mathcal{S}(\mathbf{w})$ is absolutely convergent, we need to show that the series $\mathcal{S}'(\mathbf{w})$, obtained by replacing each term of $\mathcal{S}(\mathbf{w})$ by its absolute value, is convergent. We can write

$$\begin{aligned}
\mathcal{S}'(\mathbf{w}) \;=\; \sum_{\mathbf{r}\in\mathbb{Z}_+^K} \left| \frac{\mathbf{w}^{\mathbf{r}}}{(\mathbf{r})!} \right| \cdot \left| \int \mathbf{a}^{\mathbf{r}} \; d\tilde{\Lambda}^{(t)}(\mathbf{a}) \right| \;&=\; \sum_{\mathbf{r}\in\mathbb{Z}_+^K} \frac{\mathbf{w}_+^{\mathbf{r}}}{(\mathbf{r})!} \cdot \left| \int \mathbf{a}^{\mathbf{r}} \; d\tilde{\Lambda}^{(t)}(\mathbf{a}) \right| \\
&\leq\; \sum_{\mathbf{r}\in\mathbb{Z}_+^K} \frac{\mathbf{w}_+^{\mathbf{r}}}{(\mathbf{r})!} \cdot C_{k,d,\delta} \cdot \left(\frac{t}{\delta} \right)^{|\mathbf{r}|} \\
&=\; C_{k,d,\delta} \cdot \exp\left(\frac{t}{\delta} \cdot \sum_{i=1}^{K} |w_i| \right) .
\end{aligned}$$

The last equality above used the fact that for all $\mathbf{x} \in \mathbb{R}^K$, the series $\sum_{\mathbf{r}\in\mathbb{Z}_+^K} \frac{\mathbf{x}^{\mathbf{r}}}{(\mathbf{r})!}$ converges to $\exp\left(\sum_{i=1}^K x_i\right)$. ∎

Thus, we know that for $\mathbf{w} \in \left[-\frac{\beta}{t+1}, \frac{\beta}{t+1} \right]^K$, the series $\mathcal{S}(\mathbf{w})$ always converges to zero. We shall use this to show that all the coefficients of the series must be zero, which in turn implies that the signed measure $\tilde{\Lambda}^{(t)}$ must be identically zero. The following lemma finishes the proof.

Lemma 3.19 *Let $\tilde{\Lambda}$ be a signed measure on vectors $\mathbf{a} = (a_1, \ldots, a_K)$ contained in a compact set $X \subseteq \mathbb{R}^K$, such that the series*

$$\mathcal{S}(\mathbf{w}) \;=\; \sum_{\mathbf{r}\in\mathbb{Z}_+^K} \frac{\mathbf{w}^{\mathbf{r}}}{(\mathbf{r})!} \cdot \int \mathbf{a}^{\mathbf{r}} \; d\tilde{\Lambda}(\mathbf{a})$$

in the variables $w_1, \ldots, w_K$ converges and is identically zero for $|w_i| \leq \tau$. Then $\tilde{\Lambda} = 0$.

Proof: Since the series converges for all $\mathbf{w} \in [-\tau, \tau]^K$, $\mathcal{S}(\mathbf{w})$ defines a real analytic function for all $\mathbf{w} \in [-\tau, \tau]^K$. Since the function is identically zero in $[-\tau, \tau]^K$, all its derivatives at $\mathbf{w} = (0, \ldots, 0)$ must be zero. By comparing coefficients of the above series with the Taylor expansion, we get that

$$\int \mathbf{a}^{\mathbf{r}} \; d\tilde{\Lambda}(\mathbf{a}) = 0 \qquad \forall \mathbf{r} \in \mathbb{Z}_+^K \,.$$

Thus, for all polynomials P in the variables $(a_1, \ldots, a_K)$, we have that $\int P(\mathbf{a})\ d\tilde{\Lambda}(\mathbf{a}) = 0$. By the the Stone-Weierstrass theorem, we know that for any continuous function $h : X \to \mathbb{R}$, there is a sequence of polynomials $\{P_r\}_{r\in\mathbb{N}}$, which converges to h. By the dominated convergence theorem for integrals over signed measures, we have that

$$\int h(\mathbf{a})\ d\tilde{\Lambda}(\mathbf{a}) \ = \ \lim_{r\to\infty} \int P_r(\mathbf{a})\ d\tilde{\Lambda}(\mathbf{a}) \ = \ 0\,.$$

Finally, we use the (uniqueness part of) Riesz Representation Theorem (see Chapter 13 in [37]), which says for a compact metric space X and two signed measures Λ_1 and Λ_2 defined on X, if $\int h(\mathbf{a})\ d\Lambda_1(\mathbf{a}) = \int h(\mathbf{a})\ d\Lambda_2(\mathbf{a})$ for all continuous functions $h : X \to \mathbb{R}$, then $\Lambda_1 = \Lambda_2$. Using this theorem, we conclude that $\tilde{\Lambda} = 0$. ■

The above lemma gives that the signed measure $\tilde{\Lambda}^{(t)}$ must be identically zero. Note that to apply the lemma, we use the fact that the space of the vectors $\mathbf{a}$ is compact. This follows from the fact that the space of the matrices ζ' is compact and a continuous map preserves compactness.

However, we are interested in the measure $\Lambda^{(t)}$, and we have $\tilde{\Lambda}^{(t)} = g_\beta^{(t)} \cdot \Lambda^{(t)}$. The following lemma shows that then we must in fact have that $\Lambda^{(t)} = 0$.

Lemma 3.20 *Let Λ_1 and Λ_2 be two signed measures on a compact metric space X such that $\Lambda_2 = g \cdot \Lambda_1$ for a strictly positive and bounded continuous function g. Then if Λ_2 is identically zero, so is Λ_1.*

Proof: We consider the integral of any continuous function $h : X \to \mathbb{R}$ with respect to Λ_1. Note that since g is strictly positive and X is compact, g is also bounded below by some absolute constant. Using the fact that g is positive and bounded, we can write

$$\int h\ d\Lambda_1 \ = \ \int \frac{h}{g} \cdot g\ d\Lambda_1 \ = \ \int \frac{h}{g}\ d\Lambda_2\,.$$

Since h and g are both continuous and g is positive, the function $\frac{h}{g}$ is continuous and hence measurable. Thus, we obtain that for every continuous function h,

$$\int h\ d\Lambda_1 \ = \ \int \frac{h}{g}\ d\Lambda_2 \ = \ 0\,.$$

Again, by the (uniqueness aspect of) Riesz Representation Theorem as in the proof of Lemma 3.19, this implies that $\Lambda_1 = 0$. ■

Since the function $g_\beta^{(t)}$ is strictly positive, bounded and continuous by Claim 3.15, the previous claim implies that the measure Λ on $\mathcal{C}_\delta(f)$ is such that for each $t \in [k]$, the signed measure

$$\Lambda^{(t)} \ = \ \mathop{\mathbb{E}}_{|S|=t} \ \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \ \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \left[\hat{f}(S) \cdot \left(\prod_{i=1}^{t} b_i \right) \cdot \Lambda_{S,\pi,b} \right]$$

is identically zero i.e., Λ is a vanishing measure. However, we need to establish the existence of a vanishing measure on $\mathcal{C}(f)$. The following claim shows that the existence of such measures on $\mathcal{C}(f)$ and $\mathcal{C}_\delta(f)$ are equivalent.

Claim 3.21 *There exists a vanishing probability measure Λ on $\mathcal{C}_\delta(f)$ if and only if there exists a vanishing probability measure Λ' on $\mathcal{C}(f)$.*

Proof: By definition of the body $\mathcal{C}_\delta(f)$, we have that for every $\zeta \in \mathcal{C}_\delta(f)$, the matrix

$$\zeta' \;=\; \frac{\zeta - \delta \cdot \mathbb{I}_{k+1}}{1-\delta}$$

is in $\mathcal{C}(f)$, where $\mathbb{I}_{k+1}$ denotes the $(k+1) \times (k+1)$ identity matrix. The above map defines a continuous bijection from $\mathcal{C}_\delta(f)$ to $\mathcal{C}(f)$, and thus maps measurable sets to measurable sets. Thus, we can define a measure Λ' on $\mathcal{C}(f)$ where for any measurable set $A' \subseteq \mathcal{C}(f)$, we take $\Lambda'(A') = \Lambda(A)$ for A which is the inverse image of A' under the above map. Note that if $\zeta' \in \mathcal{C}(f)$ is the image of $\zeta \in \mathcal{C}_\delta(f)$ under the above map, then we also have for any S, π and b that

$$\zeta'_{S,\pi,b} \;=\; \frac{\zeta_{S,\pi,b} - \delta \cdot \mathbb{I}_{|S|}}{1-\delta}\,.$$

Thus, we also have that for every measurable set A of $(|S|+1) \times (|S|+1)$ matrices, and its image A' that $\Lambda_{S,\pi,b}(A) = \Lambda'_{S,\pi,b}(A')$. Since $\Lambda^{(t)}$ is a linear combination of the measures $\Lambda_{S,\pi,b}$ for $|S| = t$, and $\Lambda'^{(t)}$ is an identical linear combination of measures $\Lambda'_{S,\pi,b}$, $\Lambda^{(t)}$ being identically zero implies that $\Lambda'^{(t)}$ must also be identically zero.

For the reverse direction, we consider the inverse map $\zeta = (1-\delta)\cdot\zeta' + \delta \cdot \mathbb{I}_k$, which is also continuous. The rest of the argument is the same as above. ■

3.4 The Integrality Gap Instances

We now show that for a predicate f, if there exists a vanishing probability measure Λ on $\mathcal{C}(f)$, then there exists an infinite family of $\mathsf{CSP}(f)$ instances such that the SDP has optimum value $1 - o(1)$, while the value of any integer assignment lies in $[\rho(f) - o(1), \rho(f) + o(1)]$.

First, we give a description of our instance family in the continuous setting and then sketch how to discretize it. The advantage of this description is that the soundness and completeness of the instance are far easier to analyze than in the discrete setting, while continuity properties ensure that the results translate to the discrete setting as well. To ensure the continuity of various functions defined on the matrices ζ, we will instead work with a vanishing measure Λ defined on $\mathcal{C}_\delta(f)$ (for some small $\delta > 0$) instead of $\mathcal{C}(f)$. By Claim 3.21, the existence of vanishing measures on $\mathcal{C}(f)$ and $\mathcal{C}_\delta(f)$ are equivalent.

Our set of literals will be the set of all points in $\mathbb{R}^d$, where the variable represented by the point $-\mathbf{y}$ is treated as the negation of the variable represented by the point $\mathbf{y} \in \mathbb{R}^d$. The set of constraints will be given by all k-tuples of points in $\mathbb{R}^d$. We think of the constraints being generated as follows: we pick a ζ according to Λ, choose a k-tuple of points $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$ according to $\mathcal{N}_d(\zeta)$, and impose the constraint $f(\mathbf{y}_1, \ldots, \mathbf{y}_k)$. Thus, given a k-tuple $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$, the "weight" of the constraint $f(\mathbf{y}_1, \ldots, \mathbf{y}_k)$ is $\mathbb{E}_{\zeta \sim \Lambda}\left[\gamma_{k,d}\left((\mathbf{y}_1, \ldots, \mathbf{y}_k), \zeta\right)\right]$.

We remark that while it is convenient to think of the instance as above, the set of variables in fact only corresponds to $\mathbf{y} \in H$, where H is an arbitrary half-space of $\mathbb{R}^d$, say $H = \mathbb{R}_+ \times \mathbb{R}^{d-1}$. This is because the variable $-\mathbf{y}$ is supposed to be the negation of the variable $\mathbf{y}$. This means that any "assignment" to the variables, must be an odd function on $\mathbb{R}^d$. Also, we will need to be careful of the above while constructing the SDP solution.

Soundness

Let $\psi : \mathbb{R}^d \to \{\pm 1\}$ be an odd function, which forms an assignment to the variables of our continuous $\mathsf{CSP}(f)$ instance. In the continuous setting ψ may not even be measurable but such technical issues do not occur in a discrete setting, which is our goal, and so we will ignore them for the analysis below. We show that the fraction of constraints satisfied by any such assignment is $\rho(f)$.

Lemma 3.22 *Let Φ be the instance as described above and let $\psi : \mathbb{R}^d \to \{-1, 1\}$ be any measurable odd function. Then the fraction of constraints satisfied by ψ, denoted by $\mathsf{sat}(\psi)$, is equal to $\rho(f)$.*

Proof: The objective value is given by:

$$\begin{aligned}
\mathsf{sat}(\psi) &= \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} [f(\mathbf{y}_1,\ldots,\mathbf{y}_k)] \\
&= \rho(f) + \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k\sim\mathcal{N}_d(\zeta)} \left[\sum_{\substack{S\subseteq[k]\\ S\neq\emptyset}} \hat{f}(S) \cdot \prod_{i\in S} \psi(\mathbf{y}_i) \right] \\
&= \rho(f) + \mathsf{Eval}(\Lambda, \psi) \qquad (3.7)
\end{aligned}$$

By Claim 3.11, we can write $\mathsf{Eval}(\Lambda, \psi)$ as

$$\mathsf{Eval}(\Lambda, \psi) = \sum_{t=1}^{k} \int_{(\mathbb{R}^d)^t} \theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t) \cdot \left(\prod_{i=1}^{t} \psi(\mathbf{y}_i) \right) d\mathbf{y}_1 \ldots d\mathbf{y}_t \,,$$

where the function $\theta^{(t)}$ is defined as

$$\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t) := \sum_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \mathop{\mathbb{E}}_{\zeta\sim\Lambda} \left[\hat{f}(S) \cdot \left(\prod_{i=1}^{t} b_i \right) \cdot \gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t), \zeta_{S,\pi,b}\right) \right] .$$

However, since Λ is a vanishing measure, it is easy to see that $\theta^{(t)}$ must be identically zero for each t. This follows from writing $\theta^{(t)}$ as

$$\begin{aligned}
\theta^{(t)}(\mathbf{y}_1,\ldots,\mathbf{y}_t) &= \sum_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \mathop{\mathbb{E}}_{\zeta'\sim\Lambda_{S,\pi,b}} \left[\hat{f}(S) \cdot \left(\prod_{i=1}^{t} b_i \right) \cdot \gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t), \zeta'\right) \right] \\
&= \sum_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \int \hat{f}(S) \cdot \left(\prod_{i=1}^{t} b_i \right) \cdot \gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t), \zeta'\right) d\Lambda_{S,\pi,b}(\zeta') \\
&= \int \gamma_{t,d}\left((\mathbf{y}_1,\ldots,\mathbf{y}_t), \zeta'\right) \; d\Lambda^{(t)}(\zeta') \\
&= 0 \,.
\end{aligned}$$

Hence, $\mathsf{sat}(\psi) = \rho(f)$ for every measurable and odd assignment ψ to our continuous instance. ■

Completeness

We now demonstrate an SDP solution for the continuous instance, for which the value of the objective is 1. However, it is not a valid solution to the relaxation in Figure 3, as some of the

SDP constraints will not be satisfied for each tuple of variables involved in a constraint, but only in expectation over these variables (which is also the case with the continuous Gaussian version of the Feige-Schechtman instance for MAX-CUT). However, as we discretize the instance, these constraints will be satisfied upto a small error, with high probability over the participating tuple of variables. We will be able to correct these errors later, without significantly affecting the value of the SDP solution.

To construct the SDP solution, we need to specify a vector $\mathbf{v}_{(\emptyset,\emptyset)}$, a vector $\mathbf{v}_{(\mathbf{y},b)}$ for each $\mathbf{y} \in \mathbb{R}^d$ and $b \in \{-1,1\}$, and a variable $x_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}$ for all $\mathbf{y}_1,\ldots,\mathbf{y}_k \in \mathbb{R}^d$ and $\alpha \in \{-1,1\}^k$, satisfying the conditions in Figure 3. We take the vector $\mathbf{v}_{(\emptyset,\emptyset)} = \frac{1}{\sqrt{d}} \cdot \mathbf{1}$. We shall also define the vector $\mathbf{u}_\emptyset = \mathbf{v}_{(\emptyset,\emptyset)}$ for the calculations below. For each $\mathbf{y} \in \mathbb{R}^d$, we first define the following vectors.

$$\mathbf{u}_\mathbf{y} = \frac{1}{\sqrt{d}} \cdot \mathbf{y}, \qquad \mathbf{v}_{(\mathbf{y},1)} = \frac{1}{2} \cdot (\mathbf{u}_\emptyset + \mathbf{u}_\mathbf{y}) \quad \text{and} \quad \mathbf{v}_{(\mathbf{y},-1)} = \frac{1}{2} \cdot (\mathbf{u}_\emptyset - \mathbf{u}_\mathbf{y}) \,.$$

Note that $\mathbf{v}_{(-\mathbf{y},b)} = \mathbf{v}_{(\mathbf{y},-b)}$ for any $\mathbf{y} \in \mathbb{R}^d$ and $b \in \{-1,1\}$, since $-\mathbf{y}$ is simply the negation of the variable $\mathbf{y}$.

Before describing the values of the variables $x_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}$, we mention a subtle issue. Note that we need to produce one such set of variables *for every constraint* in the CSP instance, and not just for every k-tuple of variables. This means that if for some $\mathbf{y}_1,\ldots,\mathbf{y}_k \in H$, there are two constraints of the form $f(\mathbf{y}_1,\ldots,\mathbf{y}_k)$ and $f(-\mathbf{y}_1,\ldots,\mathbf{y}_k)$, then we will produce *two different sets of variables*, $\left\{x_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}\right\}_{\alpha\in\{-1,1\}^k}$ and $\left\{x_{((-\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}\right\}_{\alpha\in\{-1,1\}^k}$, corresponding to the same tuple $(\mathbf{y}_1,\ldots,\mathbf{y}_k)$ of CSP variables. Similarly for constraints where the tuple $(\mathbf{y}_1,\ldots,\mathbf{y}_k)$ is generated according to two different matrices ζ and ζ' in the support of Λ. The only consistency conditions are the ones imposed through the inner products of the corresponding vectors.

Since every constraint is uniquely described by a tuple $(\mathbf{y}_1,\ldots,\mathbf{y}_k) \in \mathbb{R}^d$ and $\zeta \in \Lambda$, we have a different set of variables $\left\{x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}\right\}_{\alpha\in\{-1,1\}^k}$ for each ζ and $\mathbf{y}_1,\ldots,\mathbf{y}_k \in \mathbb{R}^d$. We now describe the value of the variable $x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}$ for all $\mathbf{y}_1,\ldots,\mathbf{y}_k \in \mathbb{R}^d$ and $\alpha \in \{-1,1\}^k$ and $\zeta \in \mathcal{C}_\delta(f)$. Since the only "actual variables" correspond to $\mathbf{y} \in H$, some of the elements $\mathbf{y}_i$ might be negations of actual variables $-\mathbf{y}_i \in H$. In that case we interpret $x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_i,\ldots,\mathbf{y}_k),\alpha)}$ as $x^{(\zeta)}_{((\mathbf{y}_1,\ldots,-\mathbf{y}_i,\ldots,\mathbf{y}_k),\alpha')}$ (for the constraint corresponding to $(\mathbf{y}_1,\ldots,\mathbf{y}_k)$ and ζ), where α' is α with the i^{th} bit negated.

Recall that for each $\zeta \in \mathcal{C}_\delta(f)$, there exists a distribution ν supported on $f^{-1}(1)$, such that $\zeta = (1-\delta)\cdot\zeta(\nu) + \delta \cdot \mathbb{I}_{k+1}$. Consider a distribution $\overline{\nu}$, which is ν with probability $1-\delta$ and uniform on $\{-1,1\}^k$ with probability δ. Then, we have

$$\zeta \;=\; (1-\delta)\cdot\zeta(\nu) + \delta\cdot\mathbb{I}_{k+1} \;=\; \zeta(\overline{\nu})\,.$$

We refer to (an arbitrary choice of) this distribution $\overline{\nu}$ for a given $\zeta \in \mathcal{C}_\delta(f)$ as $\overline{\nu}_\zeta$. For a constraint $f(\mathbf{y}_1,\ldots,\mathbf{y}_k)$, the variable $x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}$ is then defined as

$$x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)} \;=\; \overline{\nu}_\zeta(\alpha)\,,$$

where $\overline{\nu}_\zeta(\alpha)$ is the probability assigned to α by $\overline{\nu}_\zeta$. We now show that this assignment has SDP value $(1-\delta)$ and satisfies the SDP constraints in expectation over the tuples $(\mathbf{y}_1,\ldots,\mathbf{y}_k)$.

Lemma 3.23 *Let Φ be the continuous instance of $\mathsf{CSP}(f)$ as described above. Then the SDP solution given by the vectors $\mathbf{v}_{(\mathbf{y},b)}$ and the variables $x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}$ defined as above has an objective value of $1-\delta$. Also, we have*

- *For all $i \in [k]$ and all ζ in the support of Λ, $\mathbb{E}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\left\langle \mathbf{v}_{(\mathbf{y}_i,1)}, \mathbf{v}_{(\mathbf{y}_i,-1)} \right\rangle\right] = 0$.*
- *For all $i, j \in [k]$ with $i \neq j$, all $b, b' \in \{-1,1\}$, and all ζ in the support of Λ,*

$$\mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\left\langle \mathbf{v}_{(\mathbf{y}_i,b)}, \mathbf{v}_{(\mathbf{y}_j,b')} \right\rangle\right] \;=\; \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\sum_{\substack{\alpha \in \{-1,1\}^t \\ \alpha(i)=b, \alpha(j)=b'}} x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)} \right].$$

The remaining SDP conditions are satisfied for each constraint corresponding to a tuple $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$ and matrix ζ.

Proof: We first verify the SDP constraints. It is immediate from the definitions that we have $\left\|\mathbf{v}_{(\emptyset,\emptyset)}\right\| = 1$, $\mathbf{v}_{(\mathbf{y},1)} + \mathbf{v}_{(\mathbf{y},-1)} = \mathbf{v}_{(\emptyset,\emptyset)}$ for all $\mathbf{y} \in \mathbb{R}^d$, and $x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)} \geq 0$ for all $\mathbf{y}_1, \ldots, \mathbf{y}_k \in \mathbb{R}^d$ and all $\alpha \in \{-1,1\}^k$. The remaining two constraints will only be satisfied in expectation over the tuple $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$.
Consider the constraint $\left\langle \mathbf{v}_{(\mathbf{y},1)}, \mathbf{v}_{(\mathbf{y},-1)} \right\rangle = 0$. With our definition of vectors, we have

$$\left\langle \mathbf{v}_{(\mathbf{y},1)}, \mathbf{v}_{(\mathbf{y},-1)} \right\rangle \;=\; \frac{1}{4} \cdot \left(\|\mathbf{u}_\emptyset\|^2 - \|\mathbf{u}_\mathbf{y}\|^2 \right) \;=\; \frac{1}{4} \cdot \left(1 - \frac{1}{d} \cdot \|\mathbf{y}\|^2 \right),$$

which is not always zero. However, for any $i \in [k]$, we have

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\left\langle \mathbf{v}_{(\mathbf{y}_i,1)}, \mathbf{v}_{(\mathbf{y}_i,-1)} \right\rangle\right] \;&=\; \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\frac{1}{4} \cdot \left(1 - \frac{1}{d} \cdot \|\mathbf{y}_i\|^2 \right) \right] \\ &=\; \frac{1}{4} \cdot (1 - \zeta(i,i)) \\ &=\; 0\,. \end{aligned}$$

Thus, the constraint is satisfied in expectation over the tuples $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$ for each ζ. Similarly, for any tuple $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$, $i, j \in [k], i \neq j$ and $b, b' \in \{-1,1\}$, we have the constraint

$$\sum_{\substack{\alpha \in \{-1,1\}^t \\ \alpha(i)=b, \alpha(j)=b'}} x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)} \;=\; \left\langle \mathbf{v}_{(\mathbf{y}_i,b)}, \mathbf{v}_{(\mathbf{y}_j,b')} \right\rangle.$$

From the definition of the variables $x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}$, the left hand side equals

$$\begin{aligned} \mathop{\mathbb{P}}_{z \sim \overline{\nu}_\zeta} \left[(z_i = b) \wedge (z_j = b') \right] \;&=\; \mathop{\mathbb{E}}_{z \sim \overline{\nu}_\zeta} \left[\left(\frac{1 + (-1)^b \cdot z_i}{2} \right) \cdot \left(\frac{1 + (-1)^{b'} \cdot z_j}{2} \right) \right] \\ &=\; \frac{1}{4} \cdot \left(1 + (-1)^b \cdot \zeta(0,i) + (-1)^{b'} \cdot \zeta(0,j) + (-1)^{b+b'} \cdot \zeta(i,j) \right). \end{aligned}$$

Also, the right hand side equals

$$\begin{aligned} \left\langle \mathbf{v}_{(\mathbf{y}_i,b)}, \mathbf{v}_{(\mathbf{y}_j,b')} \right\rangle \;&=\; \left\langle \left(\frac{\mathbf{u}_\emptyset + (-1)^b \cdot \mathbf{u}_{\mathbf{y}_i}}{2} \right), \left(\frac{\mathbf{u}_\emptyset + (-1)^{b'} \cdot \mathbf{u}_{\mathbf{y}_j}}{2} \right) \right\rangle \\ &=\; \frac{1}{4} \cdot \left(1 + \frac{(-1)^b}{d} \cdot \langle \mathbf{1}, \mathbf{y}_i \rangle + \frac{(-1)^{b'}}{d} \cdot \langle \mathbf{1}, \mathbf{y}_j \rangle + \frac{(-1)^{b+b'}}{d} \cdot \langle \mathbf{y}_i, \mathbf{y}_j \rangle \right). \end{aligned}$$

Again, we have in expectation over the tuples $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\left\langle \mathbf{v}_{(\mathbf{y}_i, b)}, \mathbf{v}_{(\mathbf{y}_j, b')} \right\rangle \right] \\
&= \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\frac{1}{4} \cdot \left(1 + \frac{(-1)^b}{d} \cdot \langle \mathbf{1}, \mathbf{y}_i \rangle + \frac{(-1)^{b'}}{d} \cdot \langle \mathbf{1}, \mathbf{y}_j \rangle + \frac{(-1)^{b+b'}}{d} \cdot \langle \mathbf{y}_i, \mathbf{y}_j \rangle \right) \right] \\
&= \frac{1}{4} \cdot \left(1 + (-1)^b \cdot \zeta(0, i) + (-1)^{b'} \cdot \zeta(0, j) + (-1)^{b+b'} \cdot \zeta(i, j) \right).
\end{aligned}$$

Thus, the SDP constraint is satisfied in expectation over the tuples $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$. Finally, we verify that the above solution has an SDP value of $1 - \delta$. The expression for the SDP value can be written as

$$\begin{aligned}
\mathop{\mathbb{E}}_{\zeta \sim \Lambda} \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\sum_{\alpha \in \{-1,1\}^k} f(\alpha) \cdot x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)} \right] &= \mathop{\mathbb{E}}_{\zeta \sim \Lambda} \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} \left[\mathop{\mathbb{P}}_{\alpha \sim \overline{\nu}_\zeta} [f(\alpha) = 1] \right] \\
&\geq \mathop{\mathbb{E}}_{\zeta \sim \Lambda} \mathop{\mathbb{E}}_{\mathbf{y}_1,\ldots,\mathbf{y}_k \sim \mathcal{N}_d(\zeta)} [(1 - \delta)] \,,
\end{aligned}$$

since $\overline{\nu}_\zeta$ is a convex combination of ν with probability $1-\delta$ and uniform on $\{-1, 1\}^k$ with probability δ, and ν is supported on $f^{-1}(1)$. ∎

Discretization

We now describe how to discretize the continuous instance described above. We first discretize the body $\mathcal{C}_\delta(f)$ and replace it by a sufficiently dense set of points. The measure Λ can then be replaced by a distribution Λ' over these set of points. Recall that the value of any integer assignment ψ to the continuous instance generated according to the measure Λ is $\rho(f) + \mathsf{Eval}(\Lambda, \psi)$ as derived in Equation 3.7. Since the function $\mathsf{Eval}(\cdot, \psi)$ is continuous in the matrices ζ (for $\zeta \in \mathcal{C}_\delta(f)$) by Lemma 3.4, replacing Λ by Λ' only affects the value of the assignment ψ by $o(1)$. Hence, the value of each assignment is in $[\rho(f) - o(1), \rho(f) + o(1)]$.

Next we restrict the set of constraints. We say that a constraint on the tuple $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$ generated according to a matrix ζ is ε-good, if for all $i, j \in [k]$, we have

$$\left| \frac{1}{d} \cdot \langle \mathbf{y}_i, \mathbf{1} \rangle - \zeta(0, i) \right| \leq \varepsilon \qquad \text{and} \qquad \left| \frac{1}{d} \cdot \langle \mathbf{y}_i, \mathbf{y}_j \rangle - \zeta(i, j) \right| \leq \varepsilon \,.$$

We will restrict our set of constraints only to the set of ε-good constraints, for a sufficiently small ε to be fixed later. Since the tuple $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$ is generated according to $\mathcal{N}_d(\zeta)$, we have that $\mathbb{E}\left[\frac{1}{d} \cdot \langle \mathbf{y}_i, \mathbf{1} \rangle = \zeta(0, i)\right]$ and $\mathbb{E}\left[\frac{1}{d} \cdot \langle \mathbf{y}_i, \mathbf{y}_j \rangle\right] = \zeta(i, j)$. Hence for sufficiently large d, the probability that a randomly generated constraint is *not* ε-good is $o(1)$ by standard tail estimates on Gaussian variables. Thus, restricting our instance only to the set of ε-good constraints changes the value of all assignments only by $o(1)$. Note that it follows from the proof of Lemma 3.23 that for any ε-good constraint, we will have for all $i, j \in [k]$ and $b, b' \in \{-1, 1\}$

$$\left| \left\langle \mathbf{v}_{(\mathbf{y}_i, 1)}, \mathbf{v}_{(\mathbf{y}_i, -1)} \right\rangle \right| \leq \varepsilon \quad \text{and} \quad \left| \left\langle \mathbf{v}_{(\mathbf{y}_i, b)}, \mathbf{v}_{(\mathbf{y}_j, b')} \right\rangle - \sum_{\substack{\alpha \in \{-1,1\}^t \\ \alpha(i) = b, \alpha(j) = b'}} x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)} \right| \leq \varepsilon \,.$$

Finally, we discretize the set of variables. Since we only consider ε-good constraints, we have that for all participating tuples $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$ and all $i \in [k]$, $\left|\frac{1}{d} \cdot \langle \mathbf{y}_i, \mathbf{y}_i \rangle - \zeta(i,i)\right| \leq \varepsilon$ and hence $\|\mathbf{y}_i\|^2 \in [(1-\varepsilon)\cdot d, (1+\varepsilon)\cdot d]$. Thus, we can restrict ourselves to a sufficiently dense set of points such that their squared distance from the origin is between $(1-\varepsilon)\,d$ and $(1+\varepsilon)\,d$. For each constraint on a tuple $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$, we collapse each $\mathbf{y}_i$ to the nearest point in our set, which gives a finite set of constraints over a finite number of variables. Since an assignment to the collapsed instance can also be thought of as an assignment to the continuous instance (where ψ is constant over each set of collapsed points), the value of any assignment still remains in the range $[\rho(f) - o(1), \rho(f) + o(1)]$.

We define the vectors $\mathbf{v}_{(\mathbf{y},b)}$ and variables $x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)}$ as before for our new set of variables. Since the contribution of *each constraint* to the SDP objective is at least $1-\delta$, the SDP value still remains at least $1-\delta$. Also, if the set of points is sufficiently dense, each vector only moves by a small amount (say $o(\varepsilon)$) and we still have that for every (ε-good) constraint, for all $i, j \in [k]$ and $b, b' \in \{-1, 1\}$

$$\left|\left\langle \mathbf{v}_{(\mathbf{y}_i,1)}, \mathbf{v}_{(\mathbf{y}_i,-1)} \right\rangle\right| \leq O(\varepsilon) \quad \text{and} \quad \left| \left\langle \mathbf{v}_{(\mathbf{y}_i,b)}, \mathbf{v}_{(\mathbf{y}_j,b')} \right\rangle - \sum_{\substack{\alpha \in \{-1,1\}^t \\ \alpha(i)=b, \alpha(j)=b'}} x^{(\zeta)}_{((\mathbf{y}_1,\ldots,\mathbf{y}_k),\alpha)} \right| \leq O(\varepsilon)\,.$$

Thus, we have an SDP solution with value at least $1-\delta$, which satisfies the above inequalities approximately and the rest of the SDP constraints exactly. At this point we can apply the "surgery" and "smoothening" procedures of Raghavendra and Steurer [30] (Lemmas 5.1 and 5.2), which transform an SDP solution satisfying the above constraints approximately, to new solution for the basic SDP relaxation in Figure 3, while only losing $O(\sqrt{\varepsilon}\cdot k^2)$ in the SDP value. Note that for an instance of $\mathsf{CSP}(f)$, the variables $x_{(S_C,\alpha)}$ define a distribution on the set S_C. Let this be denoted by ν_C. The following is a combination of Lemmas 5.1 and 5.2 from [30].

Lemma 3.24 ([30]) *Let Φ be an instance of $\mathsf{CSP}(f)$ in n (Boolean) variables such that there exist vectors $\mathbf{v}_{(i,b)}$ for all $i \in [n]$ and $b \in \{-1,1\}$, and distributions ν_C over $\{-1,1\}^{S_C}$ for all $C \in \Phi$, satisfying*

$$\left|\left\langle \mathbf{v}_{(i,1)}, \mathbf{v}_{(i,-1)} \right\rangle\right| \leq \varepsilon \quad \text{and} \quad \left| \left\langle \mathbf{v}_{(i,b)}, \mathbf{v}_{(j,b')} \right\rangle - \mathop{\mathbb{P}}_{x \sim \nu_C} \left[(x_i = b) \wedge (x_j = b')\right] \right| \leq \varepsilon \quad \forall C \in \Phi,\ i,j \in S_C\,.$$

Then there exist vectors $\left\{\tilde{\mathbf{v}}_{(i,b)}\right\}_{i\in[n], b\in\{-1,1\}}$ and distributions $\{\tilde{\nu}_C\}_{C\in\Phi}$ such that

$$\left\langle \tilde{\mathbf{v}}_{(i,1)}, \tilde{\mathbf{v}}_{(i,-1)} \right\rangle = 0 \quad \text{and} \quad \left\langle \tilde{\mathbf{v}}_{(i,b)}, \tilde{\mathbf{v}}_{(j,b')} \right\rangle = \mathop{\mathbb{P}}_{x \sim \tilde{\nu}_C} \left[(x_i = b) \wedge (x_j = b')\right] \quad \forall C \in \Phi,\ i,j \in S_C\,.$$

Also, we have that for all i, b, $\left\|\mathbf{v}_{(i,b)} - \tilde{\mathbf{v}}_{(i,b)}\right\| = O(k^2 \cdot \sqrt{\varepsilon})$ and for all $C \in \Phi$, $\|\nu_C - \tilde{\nu}_C\|_1 = O(k^2 \cdot \sqrt{\varepsilon})$.

Choosing $\varepsilon = O(\delta^2/k^4)$ and applying the above lemma, we obtain a solution to the SDP in Figure 3 with value at least $1 - 2\delta$.

4 Proof of the LP Dichotomy Theorem

Note that a dichotomy theorem for SDPs need not imply a similar dichotomy theorem for LPs. For example, 2LIN is approximable (very well) via Goemans-Williamson SDP but the same predicate

appears approximation resistant to a super-constant number of rounds of the Sherali-Adams LP [13, 10]. Nevertheless, our characterization in Theorem 2.14 can be used in a more or less black-box fashion so as to yield a *syntactically* similar characterization in the LP case. The integrality gap construction however needs substantial work. The Feige-Schechtman approach is not sufficient to construct integrality gap instances for the Sherali-Adams LP, which is our focus in this section. We overcome this difficulty by generalizing the construction of de la Vega and Kenyon [13]. A noteworthy detail of our construction is that our technique, even though it is probabilistic, requires a more subtle argument for both completeness and soundness. This is unlike many previous constructions, which typically consider a uniformly random instance (or a minor modification of it) from the family of all possible instances.

Recall that in Definition 2.12, we define a moment matrix ζ consisting of the first and second moments of a distribution ν supported on $f^{-1}(1)$. The second moments also match with the inner products of the SDP vectors. In the LP case, the LP solution only gives first moments. Still, we are able to use a *dummy* setting for the second moments and reduce the LP case to the SDP case! The dummy setting ensures that the corresponding covariances are zero and hence the Gaussians with matching first and second moments are independent.

We describe this trick formally now. Given a predicate $f : \{-1,1\}^k \to \{0,1\}$, recall that $\mathcal{D}(f)$ is the set of all probability distributions over $f^{-1}(1)$. We define a compact body $\tilde{\mathcal{C}}(f)$ that replaces the role of the polytope $\mathcal{C}(f)$ before.

Definition 4.1 *For $\nu \in \mathcal{D}(f)$, we let $\tilde{\zeta}(\nu)$ denote the $(k+1) \times (k+1)$ symmetric* moment matrix*:*

$$\begin{aligned} \forall i \in \{0\} \cup [k] : \ \tilde{\zeta}(i,i) &= 1 \,, \\ \forall i \in [k] : \ \tilde{\zeta}(0,i) &= \mathop{\mathbb{E}}_{x \sim \nu}[x_i] \,, \\ \forall i,j \in [k], i \neq j : \ \tilde{\zeta}(i,j) &= \tilde{\zeta}(0,i) \cdot \tilde{\zeta}(0,j) \,. \end{aligned}$$

Also, let $\tilde{\mathcal{C}}(f) \subseteq \mathbb{R}^{(k+1)\times(k+1)}$ denote the compact (but not necessarily convex) set of all such moment matrices:

$$\tilde{\mathcal{C}}(f) \ := \ \{\tilde{\zeta}(\nu) \ : \ \nu \in \mathcal{D}(f)\}.$$

Note that if $g_1, \ldots, g_k$ are correlated Gaussians with $\mathbb{E}[g_i] = \tilde{\zeta}(0,i)$, $\mathbb{E}[g_i^2] = 1$ and $\mathbb{E}[g_i g_j] = \tilde{\zeta}(0,i) \cdot \tilde{\zeta}(0,j)$ for $i \neq j$, then these are independent with given means.

The entire argument in Section 3 can be repeated as is except for two changes: firstly, the body $\tilde{C}(f)$ is now used throughout the argument. Secondly, in Section 3.2, towards designing an algorithm, the k-round Sherali-Adams LP is solved instead of the basic (SDP) relaxation. The SDP solution enables us to generate a (global) sequence of correlated Gaussians, one for every CSP variable, so that for every CSP constraint C, the k Gaussians corresponding to that constraint have first and second moments given by $\zeta = \zeta(\nu(C))$, where $\nu(C)$ is the local distribution on that constraint. In the LP case however, we only have access to (globally consistent) first moments (i.e. biases) of the local distributions $\nu(C)$. But we can still generate a (global) sequence of correlated Gaussians as before whose first and second moments corresponding to the constraint C are $\tilde{\zeta}(\nu(C)))$. These are simply independent unit ℓ_2-norm Gaussians with first moments equal to the biases computed by the LP!

As before, depending on the value of the limit L, we get a dichotomy, i.e. the following analogs of Theorems 3.5 and 3.6 respectively. When $L > 0$, the predicate is weakly approximable via a k-round Sherali-Adams LP.

Theorem 4.2 *If $L > 0$, then there exists a k-round LP rounding algorithm such that given an instance Φ with $\mathsf{FRAC}(\Phi) \geq 1-\varepsilon$ (for sufficiently small $\varepsilon > 0$), we have $\mathbb{E}_{\psi} |\mathsf{ROUND}_{\psi}(\Phi) - \rho(f)| \geq L/2$.*

When $L = 0$, as in Theorem 3.6, we get a measure Λ on the body $\tilde{C}(f)$ that is vanishing in the sense therein. However we note that since the second moments are just dummy, we might as well restrict everything to the body $\mathcal{C}^*(f)$ that is the projection of $\tilde{C}(f)$ onto the first moments (and thus $\mathcal{C}^*(f)$ is simply the convex hull of $f^{-1}(1)$). Denoting the measure on $\mathcal{C}^*(f)$ so obtained by Λ^*, we get:

Theorem 4.3 *If $L = 0$, then there exists a probability measure Λ^* on $\mathcal{C}^*(f)$ such that for all $t \in [k]$, and a uniformly random choice of S with $|S| = t$, $\pi : [t] \to [t]$ and $b \in \{-1,1\}^t$, the following signed measure on t-dimensional vectors,*

$$\Lambda^{*,(t)} := \mathop{\mathbb{E}}_{|S|=t} \mathop{\mathbb{E}}_{\pi:[t]\to[t]} \mathop{\mathbb{E}}_{b\in\{-1,1\}^t} \left[\hat{f}(S) \cdot \left(\prod_{i=1}^{t} b_i \right) \cdot \Lambda^*_{S,\pi,b} \right] \tag{4.1}$$

is identically zero.

In the next section, we show how the existence of the measure Λ^* leads to a strong $(1-o(1), \rho(f) \pm o(1))$ integrality gap for a super-constant number of rounds of the Sherali-Adams LP.

4.1 The Integrality Gap Instance

The integrality gap construction for the Sherali-Adams LP is rather different from that for SDPs. One important aspect of our construction is that unlike many previous constructions, e.g. [10, 13, 35, 8], our construction requires a non-trivial proof of *both* the soundness and completeness parts. The proof of the soundness part is similar to that in the SDP case and for the completeness part we generalize the construction in de la Vega and Kenyon [13]. A formal description of our instance follows.

Let $f : \{-1,1\}^k \mapsto \{0,1\}$ be any predicate with a measure Λ^* as in Theorem 4.3. Note that now $\mathcal{C}^*_\delta(f)$ is simply the body $\{(1-\delta)\cdot\zeta \mid \zeta \in \mathcal{C}^*(f)\}$. Since it's just a scaling, there is a vanishing measure over $\mathcal{C}^*(f)$ if and only if there is such a measure over $\mathcal{C}^*_\delta(f)$. We will assume Λ^* is over $\mathcal{C}^*_\delta(f)$ for reasons similar to the ones in the SDP integrality gap. We will finally need to choose $\delta \geq \sqrt{\varepsilon}$, for the constant ε below.

Fix a small enough $\varepsilon > 0$ and let $s = \lceil \frac{1}{\varepsilon} \rceil$. Partition the interval $[0,1]$ into $s+1$ disjoint sets $I_0, I_1, \ldots, I_s$ where $I_0 = \{0\}$ and $I_1, \ldots, I_s$ are contiguous equal length intervals that partition $(0,1]$, each being open at its left endpoint and closed at the right endpoint. For each interval I_i, we define a set (layer) of n variables X_i. Thus the total number of variables in the CSP instance is $(s+1)\cdot n$. Our constraints are generated by the following algorithm.

1. Sample[12] $\zeta \sim \Lambda^*$.
2. For each $j \in [k]$, let $i_j(\zeta)$ denote the index of the interval that contains $|\zeta(j)|$. Sample uniformly a variable x_{i_j} from the set $X_{i_j(\zeta)}$.

[12]Strictly speaking, it is not clear how to sample from an arbitrary measure Λ^*. However, as in the case of the SDP integrality gap instance, we can approximate the body $\mathcal{C}^*_\delta(f)$ by a sufficiently dense set of points, and consider Λ^* to be a distribution over a finite set of points.

3. If $\zeta(j) < 0$ then negate x_{i_j}. If $\zeta(j) = 0$ then negate x_{i_j} with probability $\frac{1}{2}$.

4. We have sampled a k-tuple of literals. Introduce a constraint f on these literals.

5. Repeat the above procedure $m = \Delta(\varepsilon) \cdot n$ times independently (where $\Delta(\varepsilon)$ is a sufficiently large constant) and thus generate m constraints.

This completes the description of our CSP(f) instance.

Let ψ be any (global) $\{-1, 1\}$-assignment to the above instance. Denoting the fraction of constraints satisfied by ψ by $\mathsf{sat}(\psi)$, we note that $\mathbb{E}[\mathsf{sat}(\psi)]$ is equal to the the probability that a randomly chosen constraint as above is satisfied by ψ. We prove that this probability is precisely $\rho(f)$. We can write $\mathbb{E}[\mathsf{sat}(\psi)]$, which equals the probability of satisfying a random constraint as above, as:

$$\mathbb{E}[\mathsf{sat}(\psi)] \quad = \quad \mathop{\mathbb{E}}_{\substack{\zeta \sim \Lambda^*, \\ x_{i_j} \in X_{i_j(\zeta)}}} [f(\mathsf{sign}(\zeta(1)) \cdot \psi(x_{i_1}), ..., \mathsf{sign}(\zeta(k)) \cdot \psi(x_{i_k}))] .$$

Here the function $\mathsf{sign}(\cdot)$ is -1 if its argument is strictly negative, $+1$ if its argument is strictly positive and $\mathsf{sign}(0) = 0$. Using the Fourier expansion of f,

$$\mathbb{E}[\mathsf{sat}(\psi)] \quad = \quad \mathop{\mathbb{E}}_{\substack{\zeta \sim \Lambda^*, \\ x_{i_j} \in X_{i_j(\zeta)}}} \left[\sum_{S \subseteq [k]} \hat{f}(S) \prod_{j \in S} \left(\mathsf{sign}(\zeta(j)) \cdot \psi(x_{i_j}) \right) \right] .$$

Since x_{i_j} is randomly chosen from the layer $X_{i_j(\zeta)}$, we can move the expectation over the choice of x_{i_j} inside and get

$$\mathbb{E}[\mathsf{sat}(\psi)] \quad = \quad \rho(f) + \sum_{t=1}^{k} \mathop{\mathbb{E}}_{\zeta \sim \Lambda^*} \left[\sum_{|S|=t} \hat{f}(S) \prod_{j \in S} \left(\mathsf{sign}(\zeta(j)) \cdot \mathop{\mathbb{E}}_{x_{i_j} \in X_{i_j(\zeta)}} [\psi(x_{i_j})] \right) \right] .$$

The expectations inside are the average values of ψ over the respective layers and hence in $[-1, 1]$. Define a function $\tilde{\psi} : [-1, 1] \mapsto [-1, 1]$ that is odd, in particular $\tilde{\psi}(0) = 0$ and for each $i \in [s]$, is constant on the interval I_i where it takes the value $\mathbb{E}_{x_i \in X_i}[\psi(x_i)]$. Thus the innermost expectation is really $\tilde{\psi}(|\zeta(j)|)$ and combining it with $\mathsf{sign}(\zeta(j))$ and using the oddness of $\tilde{\psi}$,

$$\mathbb{E}[\mathsf{sat}(\psi)] \quad = \quad \rho(f) + \sum_{t=1}^{k} \mathop{\mathbb{E}}_{\zeta \sim \Lambda^*} \left[\sum_{|S|=t} \hat{f}(S) \prod_{j \in S} \tilde{\psi}(\zeta(j)) \right] . \tag{4.2}$$

We observe that for every $t \in [k]$, the expectation above vanishes. This is because, up to a multiplicative factor of $\binom{k}{t}$, the expectation is same as

$$\mathop{\mathbb{E}}_{\zeta \sim \Lambda^*} \left[\mathop{\mathbb{E}}_{|S|=t} \mathop{\mathbb{E}}_{\pi : S \mapsto S} \mathop{\mathbb{E}}_{b \in \{-1,1\}^S} \left[\hat{f}(S) \left(\prod_{j \in S} b_j \right) \left(\prod_{j \in S} \tilde{\psi}(b_j \zeta(\pi(j))) \right) \right] \right] ,$$

which in turn is same as

$$\int \left(\prod_{j=1}^{t} \tilde{\psi}(\zeta'(j)) \right) d\Lambda^{*,(t)}(\zeta').$$

This integral vanishes since $\Lambda^{*,(t)}$ vanishes identically and we are done.

Now we prove the soundness property of the CSP instance. Since each constraint is picked independently, a Chernoff bound implies that the probability that $\mathsf{sat}(\psi)$ is outside $[\rho(f)-\varepsilon, \rho(f)+\varepsilon]$, for any fixed $\{-1,1\}$ assignment ψ, decays exponentially in m. For large enough $\Delta(\varepsilon)$, one may then take a union bound over all $2^{(s+1)\cdot n}$ assignments and obtain the following claim.

Lemma 4.4 *For every $\varepsilon > 0$, there exists a sufficiently large constant $\Delta(\varepsilon)$ such that w.h.p. over the choice of the $\mathsf{CSP}(f)$ instance, it holds that for every assignment ψ to the instance, $\mathsf{sat}(\psi) \in [\rho(f)-\varepsilon, \rho(f)+\varepsilon]$.*

Let G denote the natural constraint vs variable bipartite graph of our instance. In other words, G has a vertex for each constraint and each variable and there is an edge between a constraint and a variable if and only if the variable occurs in that constraint. Strictly speaking, G is a multi-graph since in a constraint, the same variable may appear twice or more. We show that after deleting a small fraction of vertices, G has high girth, in particular eliminating cycles of length two, i.e. multiple edges.

Lemma 4.5 *The constraint vs variable graph G has $(k\Delta)^{O(g)}$ cycles of length at most g, in expectation.*

Proof: Recall that the variable vertices of G correspond to the set $[n] \times \{0, 1, \ldots, s\}$. We think of these as arranged in an $n \times (s+1)$ array. Suppose we contract the set of $s+1$ vertices in j^{th} row into a single vertex x_j for $j \in [n]$. We will get a bipartite multi-graph G' such that the set of variables of each of the m constraints is picked uniformly from the set of variables $\{x_j : j \in [n]\}$. Note that under this operation there exists a unique cycle of length at most g in G' for every cycle of length at most g in G. Moreover, the probability of obtaining that cycle in G' is the at most the probability of obtaining that cycle in G. Hence, it will suffice to bound the expected number of cycles of length at most g in G'. We have reduced our problem to obtaining a bound on the girth of G to the following combinatorial problem.

We have a random bipartite multi-graph $H := (U, V)$, where the edge set $E(H)$ is selected by independent sampling (with repetition) of k vertices from V ($|V| = n$), for each of the m vertices in U. We need a bound on the expected number of cycles of length at most g.

Consider any cycle $C(h)$ of length $2h$ in H. Half the vertices in $C(h)$ come from U and half come from V. The probability that a given vertex in U and given vertex in V have an edge between them is at most k/n. Therefore, the expected number of cycles of length exactly $2h$ in H is bounded by:

$$n^h \cdot (\Delta \cdot n)^h \cdot \left(\frac{k}{n}\right)^{2h} \leq (k\Delta)^{O(h)}. \tag{4.3}$$

The above is a geometric progression in h, since k and Δ are constants. Hence, the expected number of cycles of length at most $2h$ in H is also bounded by $(k\Delta)^{O(h)}$. ∎

For $g = c \cdot \log n$ for a sufficiently small constant c depending on k and Δ, we may delete $o(n)$ constraints from our instance so as to eliminate all cycles of length at most g. This still preserves the property that for every assignment ψ to the instance $\mathsf{sat}(\psi) \in [\rho(f)-\varepsilon, \rho(f)+\varepsilon]$, possibly with a negligible change in parameter ε that we ignore. Moreover, a union bound implies that with high probability every vertex in our constraint bigraph G has bounded degree. Therefore, Lemmas 4.4 and 4.5 imply the following lemma.

Lemma 4.6 *For all large enough n and every $\varepsilon > 0$, there exists a $\mathsf{CSP}(f)$ instance with n variables and $m = \Delta n$ constraints such that its constraint vs variable graph G has girth $\Omega(\log n)$, every vertex in G has bounded degree and every assignment to the instance satisfies between $[\rho(f) - \varepsilon, \rho(f) + \varepsilon]$ fraction of the constraints.*

Also note that large girth in particular implies that any two constraints in our instance share at most one variable.

For the remainder of this section, we assume that our $\mathsf{CSP}(f)$ instance is given by some fixed constraint graph G, as in Lemma 4.6. Next, we need to show that the Sherali-Adams LP has an optimal solution with value $1 - o(1)$ for instance given by G. Our task is to define locally consistent distributions over all subsets of variables of size at most r (we will finally be able to choose $r = \Omega(\log \log n)$). To this end we will first define distributions which are approximately consistent, and then use a result by Raghavendra and Steurer [31] to make the distributions exactly consistent.

Recall that every constraint C in our instance was generated using a $\zeta(C) \in \mathcal{C}^*(f)$. Let $\overline{\nu}(C)$ be a distribution on $f^{-1}(1)$ such that $\zeta(C) = \zeta(\overline{\nu}(C))$. Note that $\overline{\nu}(C)$ is a distribution on the *literals* involved in constraint C, with the biases of the literals being $(\zeta(1), \ldots, \zeta(k))$. If a constraint C is on variables in layers $i_1, \ldots, i_k$ respectively, then the biases of these variables according to $\overline{\nu}(C)$ are

$$(|\zeta(1)|, \ldots, |\zeta(k)|) \;=\; (p_{i_1}, \ldots, p_{i_k})\,.$$

respectively so that $p_{i_j} \in I_{i_j}$. The biases of the variables are always non-negative since we negate the j^{th} variable only if $\zeta(j) < 0$ (and with probability $1/2$ when $\zeta(j) = 0$).

The local distributions we define on sets of size r will have the property that all variables in the same layer X_i have the same bias. For each interval I_i with $i \in \{0, \ldots, s\}$, choose an arbitrary point $t_i \in I_i$. We will first modify the distributions $\overline{\nu}(C)$ such such that the all the variables in layer X_i have bias exactly t_i. Since $p_{i_j} \in I_{i_j}$, we have $|p_{i_j} - t_{i_j}| \leq \varepsilon$. Thus we can change the biases of the variables as desired with a slight perturbation of the distributions $\overline{\nu}(C)$. However this incurs a slight loss in the completeness parameter: the resulting distribution $\overline{\nu}'(C)$ is now only $(1 - o(1))$-supported on $f^{-1}(1)$.

Claim 4.7 *Let the distribution $\overline{\nu}(C)$ be as above such that the biases for the literals in C are given by $(\zeta(1), \ldots, \zeta(k))$. Also, let $t_{i_1}, \ldots, t_{i_k}$ as above be the desired biases for the variables such that $\left|t_{i_j} - |\zeta(j)|\right| \leq \varepsilon$. Then there exists a distribution $\overline{\nu}'(C)$ on $\{-1, 1\}^k$ such that $\|\overline{\nu}(C) - \overline{\nu}'(C)\|_1 = O(k \cdot \sqrt{\varepsilon})$ and*

$$\forall j \in [k] \qquad \mathop{\mathbb{E}}_{z \sim \overline{\nu}'(C)} [z_j] = \mathsf{sign}(\zeta(j)) \cdot t_{i_j}\,.$$

Thus, the biases for the variables, when the literals are sampled according to $\overline{\nu}'(C)$ are exactly $(t_{i_1}, \ldots, t_{i_k})$ since the j^{th} variable is negated only if $\mathsf{sign}(\zeta(j)) = -1$.

Proof: Let $r_j = \mathsf{sign}(\zeta(j)) \cdot t_{i_j}$ be the desired bias of the j^{th} literal. Then, $|\zeta(j) - r_j| \leq \varepsilon$ for all $j \in [k]$ We construct a sequence of distributions $\overline{\nu}_0, \ldots, \overline{\nu}_k$ such that $\overline{\nu}_0 = \overline{\nu}(C)$ and $\overline{\nu}_k = \overline{\nu}'(C)$. In $\overline{\nu}_j$, the biases of the literals are $(r_1, \ldots, r_j, \zeta(j+1), \ldots, \zeta(k))$.

The biases in $\overline{\nu}_0$ satisfy the above by definition. We think of the distributions over $z \in \{-1, 1\}^k$. We obtain $\overline{\nu}_j$ from $\overline{\nu}_{j-1}$ as,

$$\overline{\nu}_j = (1 - \tau_j) \cdot \overline{\nu}_{j-1} + \tau_j \cdot D_j\,,$$

where D_j is the distribution in which all bits, except for the j^{th} one, are set independently according to their biases in $\overline{\nu}_{j-1}$. For the j^{th} bit, we set it to $\mathsf{sign}(r_j - \zeta(j))$ (if $r_j - \zeta(j) = 0$, we can simply

proceed with $\overline{\nu}_j = \overline{\nu}_{j-1}$). The biases for all except for the j^{th} bit are unchanged. For the j^{th} bit, the bias now becomes r_j if

$$r_j = (1-\tau_j)\cdot\zeta(j) + \tau_j \cdot \mathsf{sign}(r_j - \zeta(j)) \implies \tau_j \cdot (\mathsf{sign}(r_j - \zeta(j)) - r_j) = (1-\tau_j)\cdot(r_j - \zeta(j))\,.$$

Since $\zeta \in \mathcal{C}^*_\delta(f)$ for $\delta \geq \sqrt{\varepsilon}$, we know that $|\mathsf{sign}(r_j - \zeta(j)) - r_j| \geq O(\sqrt{\varepsilon})$. Also, $|r_j - \zeta(j))| \leq \varepsilon$ by assumption. Thus, we can choose $\tau_j = O(\sqrt{\varepsilon})$ which gives that $\|\overline{\nu}_j - \overline{\nu}_{j-1}\|_1 = O(\sqrt{\varepsilon})$. The final bound then follows by triangle inequality. ■

The distribution over the literals of C, given by the above claim also gives a distribution for the variables in S_C. We now refer to the distribution over $\{-1,1\}^{S_C}$ given by Claim 4.7 as $\nu(C)$. We will need to modify the distributions $\nu(C)$ a little further before we use them to define the local distributions over sets of size r.

Definition 4.8 *Given a constraint C and $\eta > 0$, let U_C denote the following distribution on $\{-1,1\}^{S_C}$*

$$U_C \;:=\; (1-\eta)\cdot\nu(C) + \eta \cdot U_k\,.$$

where U_k denotes the uniform distribution on $\{-1,1\}^k$. For α a partial assignment to variables in C, let $U_{C,\alpha}$ denote the distribution U_C conditioned according to α.

Recall that the distributions $\nu(C)$ are defined so that the variables in the layer X_i have bias exactly t_i. The following observation will be extremely useful.

Remark 4.9 *The bias of a variable in layer X_i is exactly $(1-\eta)\cdot t_i$, when assigned according to U_C, for any constraint C containing that variable.*

Let $\mathcal{V}_G$ denote the set of variable vertices in the bipartite constraint-variable graph G and let $\mathcal{C}_G$ be the set of constraint vertices. Let $\mathsf{dist}_G(u,v)$ denote the shortest path distance in G between two vertices u and v. Given a set S of variables in G and an *even* number $d \in \mathbb{N}$, we define

$$B^{(d)}(S) \;:=\; \{u \in \mathcal{V}_G \cup \mathcal{C}_G \;:\; \mathsf{dist}_G(u,S) \leq d\}\,.$$

We will choose d to be sufficiently small so that $|B^{(d)}(S)| \leq girth(G)$ and hence the set $B^{(d)}(S)$ is a forest. Also, since d is assumed to be even and $S \subseteq \mathcal{V}_G$, the leaves of each component in $B^{(d)}(S)$ are variable vertices in G. Let $B^{(d)}(S) = \cup_i B^{(d)}(S_i)$, where each $B^{(d)}(S_i)$ is a maximal connected component in $B^{(d)}(S)$. We now describe a probabilistic process, which will be used to defined a probability distribution m_S on ± 1 assignments to the set S. We will use this process to generate a random assignment to all the variables in $B^{(d)}(S)$, and hence also in S.

First, we fix an arbitrary ordering of all variables in G. This also gives an ordering of all the constraints in S (depending on the variables involved in each constraint). We generate an assignment for all variables in $\mathcal{V}_G \cap B^{(d)}(S)$. The assignment for each component $B^{(d)}(S_i)$ is generated independently of the other components by the following process:

1. Pick the least variable $x \in S \cap B^{(d)}(S_i)$. If x belongs to the layer X_j, assign it to be 1 with probability $(1 + (1-\eta)\cdot t_j)/2$ and -1 with probability $(1 - (1-\eta)\cdot t_j)/2$, so that the bias is $(1-\eta)\cdot t_j$.
2. Traverse $B^{(d)}(S_i)$ in a breadth-first manner, starting from the vertex corresponding to the least variable x (and using the above ordering on variables and constraints).

- When visiting a vertex corresponding to a constraint C, if α is the partial assignment to the variables assigned so far, generate an assignment for the remaining variables in C according to $U_{C,\alpha}$.
- When visiting a vertex corresponding to a variable, its value is already assigned by its parent constraint-vertex. We simply proceed to its children, which are new constraint vertices.

Note that since $B^{(d)}(S_i)$ is a tree, when visiting a constraint vertex C we will have *at most one* of the variables in C assigned before. We will assign the remaining variables according to U_C conditioned on the value of this one variable.

This process above defines a probability distribution m_S on the ± 1 assignments to the variables in $B(S)$, and hence also on $\{-1,1\}^S$ as long as $B^{(d)}(S)$ is a forest. We can obtain a bound on the size of such sets S in terms of the girth and the degree of the constraint graph G.

Claim 4.10 *Let the girth of the constraint graph G be equal to g and let the degree of every vertex in G be at most D. Then the distribution m_S is well-defined for all sets S with $|S| < g/D^d$.*

Proof: Since the degree of every vertex at most D, we have that

$$\left|B^{(d)}(S)\right| \;\leq\; |S| \cdot D^d \;<\; g\,.$$

Hence, we have that $B^{(d)}(S)$ is a forest and the distribution m_S is well-defined. ■

We need the following lemma to show that the objective value of our Sherali-Adams LP solution is close to 1.

Lemma 4.11 *For every constraint C supported on variables S_C, the distribution m_{S_C}, has at least $(1-\eta-O(k\sqrt{\varepsilon}))$-fraction of its probability mass on the accepting assignments of C.*

Proof: Note that for any constraint C at most 1 variable can be fixed by a partial assignment to some other variables by our process for generating assignments. At this point, we assign all variables in C according to U_C conditioned on the value of this one variable. Hence, the joint distribution of all the variables in C is always according to U_C.

Also, U_C is obtained by taking $\nu(C)$ with probability $1-\eta$ and uniform with probability η. By Claim 4.7, $\nu(C)$ is $O(k\sqrt{\varepsilon})$-close to a distribution which corresponds to a point in $\mathcal{C}^*_\delta(f)$ and has mass at least $1-\delta$ over accepting assignments. Thus, U_C has mass at least $1-\eta-\delta-O(k\sqrt{\varepsilon})$ on accepting assignments. Using $\delta = \sqrt{\varepsilon}$ proves the bound. ■

Note that the definition of m_S implicitly depends on the ordering of variables. The following lemma shows that the distributions in fact *do not* depend on the ordering.

Lemma 4.12 *Given a set $S \subseteq \mathcal{V}_G$ and an ordering ω of all the variables in $\mathcal{V}_G$, let $m_{S,\omega}$ denote the distribution m_S when defined according the ordering ω. Then, for any $\alpha \in \{-1,1\}^S$ and any two orderings ω and ω', we have that*

$$m_{S,\omega}(\alpha) \;=\; m_{S,\omega'}(\alpha)\,.$$

Proof: Since the distributions in different components of $B^{(d)}(S)$ are independent, it is sufficient to prove the lemma for the case when $B^{(d)}(S)$ is a tree (instead of a forest). We will, in fact, prove that the probability for any assignment $\beta \in \{-1,1\}^{\mathcal{V}_G \cap B^{(d)}(S)}$ is the same regardless of the ordering ω. Since $S \subseteq \mathcal{V}_G \cap B^{(d)}(S)$, this implies the lemma.

Let $m_{S,\omega}(\beta)$ denote the probability of the assignment $\beta \in \{-1,1\}^{\mathcal{V}_G \cap B^{(d)}(S)}$. Note that since the leaves of $B^{(d)}(S)$ must correspond to variables (since d is even), for each constraint $C \in B^{(d)}(S)$, we must have that $S_C \subseteq \mathcal{V}_G \cap B^{(d)}(S)$, where S_C denotes the set of variables involved in the constraint C. For $C \in B^{(d)}(S)$, let $\beta_{|C}$ denote β restricted the set S_C.

We now compute the probability for the assignment β. Suppose that at some intermediate step in the breadth first traversal for $m_{S,\omega}$ one has fixed an assignment $\beta' \in \{\pm 1\}^R$ for a set $R \subseteq \mathcal{V}_G \cap B^{(d)}(S)$, where $\beta_{|R} = \beta'$. Let C be the next constraint-vertex visited by the traversal. Using $\beta_1 \circ \beta_2$ to denote the concatenation of two assignments β_1 and β_2, we have

$$m_{S,\omega}(\beta' \circ \beta_{|C}) \;=\; m_{S,\omega}(\beta') \cdot U_{C,\beta'}(\beta_{|C})\,,$$

where $U_{C,\beta'}(\beta_{|C})$ is the probability that constraint C gets an assignment $\beta_{|C}$ conditioned on the event that variables in R were assigned according to β'.

Since $B(S)$ is a tree, there is exactly one variable in R, say x_j, which is also present in C (this variable is the parent vertex of C). We can then write the above as

$$m_{S,\omega}(\beta' \circ \beta_{|C}) \;=\; m_{S,\omega}(\beta') \cdot \frac{U_C(\beta_{|C})}{U_C(\beta_{|j})}\,.$$

By Remark 4.9, the quantity $U_C(\beta_{|j})$ is independent of the constraint C and only depends on the variable x_j and the assignment $\beta_{|j}$. Denoting the quantity by $p_j(\beta)$, we can write the above expression as

$$m_{S,\omega}(\beta' \circ \beta_{|C}) \;=\; m_{S,\omega}(\beta') \cdot \frac{U_C(\beta_{|C})}{p_j(\beta)}\,.$$

We can now inductively simplify the expression for $m_{S,\omega}(\beta)$. Let x_{j_0} be the first variable in S according to the ordering ω. Since we visit each constraint exactly once, the numerator equals

$$p_{j_0}(\beta) \cdot \prod_{C \in \mathcal{C}_G \cap B^{(d)}(S)} U_C(\beta_{|C})\,.$$

Also, each variable $x_j \in \mathcal{V}_G \cap B^{(d)}(S)$, except for x_{j_0}, has exactly $\mathsf{deg}(x_j) - 1$ children in the tree (where $\mathsf{deg}(x_j)$ denotes its degree in the tree $B^{(d)}(S)$). Thus, the term $p_j(\beta)$ appears exactly $\mathsf{deg}(x_j) - 1$ times in the denominator, for each $x_j \in \mathcal{V}_G \cap B^{(d)}(S) \setminus \{x_{j_0}\}$. The term $p_{j_0}(\beta)$ appears $\mathsf{deg}(x_{j_0})$ times since all the neighbors of x_{j_0} are its children in the tree. Thus, we get

$$m_{S,\omega}(\beta) \;=\; \frac{\prod_{C \in \mathcal{C}_G \cap B^{(d)}(S)} U_C(\beta_{|C})}{\prod_{x_j \in \mathcal{V}_G \cap B^{(d)}(S)} (p_j(\beta))^{\mathsf{deg}(x_j)-1}}\,,$$

which is independent of the ordering ω. ∎

We now prove that the distributions m_S are *locally consistent* i.e., for any two sets S_1 and S_2, the distributions m_{S_1} and m_{S_2} agree on $S_1 \cap S_2$. It suffices to show that for that for any two sets S and T, with $S \subseteq T$, we have for all any $\alpha \in \{-1,1\}^S$, $m_S(\alpha) = m_T(\alpha)$. Here $m_T(\alpha)$ denotes the

probability that the variables in S are assigned according to α in m_T when we marginalize over the variables in $T \setminus S$. The distributions m_S will only satisfy this approximately i.e., we will be able to show that $|m_S(\alpha) - m_T(\alpha)|$ is very small. However, using a result of Raghavendra and Steurer [31], we will be able to correct the distributions $\{m_S\}$ to a family of distributions $\{m'_S\}$ such that $m'_S(\alpha) = m'_T(\alpha)$ for all α. We first prove the following.

Lemma 4.13 (Approximate Local Consistency) *There exists a constant c_0 such that for any two sets $S \subseteq T \subseteq \mathcal{V}_G$, with $|S| \leq |T| \leq 2^{c_0 \cdot \eta d}$, we have*

$$\forall \alpha \in \{-1,1\}^S \qquad |m_S(\alpha) - m_T(\alpha)| \;=\; 2^{-\Omega(\eta d)}\,,$$

when the distributions m_S and m_T are both well-defined.

Proof: Note that it suffices to prove the above for the case when $T = S \cup \{v\}$, since then by triangle inequality we will have that for any T, $|m_S(\alpha) - m_T(\alpha)| \leq |T \setminus S| \cdot 2^{-\Omega(\eta d)} = 2^{-\Omega(\eta d)}$.

Since m_T is well defined, $B^{(d)}(T)$ must be a forest in the graph G. Also, the distributions in different components of $B^{(d)}(T)$ are independent and the components in $B^{(d)}(T)$ which do not contain v are identical in $B^{(d)}(S)$. Hence, the distribution over them would be identical according to m_S and m_T. Thus, it suffices to consider the case when $B^{(d)}(T)$ is a tree (i.e., we restrict ourselves to the component $B^{(d)}(T_i)$ of $B^{(d)}(T)$ which contains v).

Note that even though $B^{(d)}(T)$ is assumed to be a tree, we could still have that $B^{(d)}(S)$ is a forest with more than one components, which get connected in $B^{(d)}(T) = B^{(d)}(S \cup \{v\})$. We first consider a simple special case when $B^{(d)}(S)$ *is also a tree.*

Case 1: $B^{(d)}(T)$ is a tree and $B^{(d)}(S)$ is also a tree. In this case, since $B^{(d)}(S) \subseteq B^{(d)}(T)$, we must have that any edge (u, v) which is present in $B^{(d)}(S)$ must also be present in $B^{(d)}(T)$. Thus, the vertices in $B^{(d)}(T) \setminus B^{(d)}(S)$ must form a collection of subtrees of the tree $B^{(d)}(S)$. By Lemma 4.12, we can assume that the distributions m_T and m_S are defined with the same starting vertex in S. Since the distribution m_T is defined by a breadth-first traversal of the tree $B^{(d)}(T)$, the probability of any assignment to the vertices in $B^{(d)}(S)$ will remain unchanged even if we remove the subtrees corresponding to the vertices in $B^{(d)}(T) \setminus B^{(d)}(S)$. Thus, in this case, we have that

$$\forall \alpha \in \{-1,1\}^S \quad m_S(\alpha) \;=\; m_T(\alpha)\,.$$

Case 2: $B^{(d)}(T)$ is a tree but $B^{(d)}(S)$ is a forest with more than one components. In this case, we might have $B^{(d)}(S) = \cup_{i=1}^{t} B^{(d)}(S_i)$, where the components $B^{(d)}(S_i)$ are disconnected in $B^{(d)}(S)$, but become connected in $B^{(d)}(T)$. Thus, the distributions over the different components will be independent according m_S but will become correlated when we consider m_T.

However, recall that in the distribution U_C, with probability η we assign the variables in S_C according to the uniform distribution on $\{-1,1\}^{S_C}$. This breaks the correlation between any two variables in the constraint C. Since for any $i \neq j$, $B^{(d)}(S_i)$ and $B^{(d)}(S_j)$ are disconnected, any path between S_i and S_j in $B^{(d)}(T)$ must have length at least d. We will use this to show that the correlation between the variables in S_i and S_j must be small since (with high probability) at some constraint C along the path, we must assign the variables in C uniformly.

However, the above intuition is slightly incorrect since in defining the distributions m_S, we do not assign all the variables of the constraint together, but assign $k-1$ variables conditioned on one variable which is the parent of C in the tree. The following claim shows that even when assigning the

variables in a constraint C, conditioned on the value of one of its variables, we break the correlation between the variables with probability at least $\eta/2$ i.e., even the conditional distribution can be viewed as being a convex combination of the uniform distribution and some other distribution.

Claim 4.14 *Let C be a constraint and let $j \in S_C$ be the index of a variable involved in C. Let $\beta \in \{-1,1\}$ be an assignment to x_j. Then the distribution $U_{C,\beta}$ on $\{-1,1\}^{S_C \setminus \{j\}}$ can be written as*

$$U_{C,\beta} \;=\; \left(1-\frac{\eta}{2}\right)\cdot m_C^{(\beta)} \;+\; \left(\frac{\eta}{2}\right)\cdot U_{k-1}\,,$$

where $m_C^{(\beta)}$ is a distribution on $\{-1,1\}^{S_C\setminus\{j\}}$ that depends on β and U_{k-1} denotes the uniform distribution on $\{-1,1\}^{S_C\setminus\{j\}}$.

Proof: Let p_β denote the probability that x is assigned the value β according to the distribution U_C. For any assignment $\beta' \in \{-1,1\}^{S_C\setminus\{j\}}$, we can write

$$\begin{aligned} U_{C,\beta}(\beta') \;=\; \frac{U_C(\beta\circ\beta')}{p_\beta} \;&=\; \frac{(1-\eta)\cdot\nu_C(\beta\circ\beta')+\eta\cdot 2^{-k}}{p_\beta} \\ &=\; \frac{(1-\eta)\cdot\nu_C(\beta\circ\beta')+\eta\cdot 2^{-k}}{p_\beta} - \eta\cdot 2^{-k} + \left(\frac{\eta}{2}\right)\cdot 2^{k-1} \\ &=\; \frac{(1-\eta)\cdot\nu_C(\beta\circ\beta')+\eta\cdot(1-p_\beta)\cdot 2^{-k}}{p_\beta} + \left(\frac{\eta}{2}\right)\cdot U_{k-1}(\beta')\,. \end{aligned}$$

Let $T_\beta(\beta')$ to denote the first term above. We can say that $T_\beta(\beta') = (1-\eta/2)\cdot m_C^{(\beta)}(\beta')$ for some distribution m_β if $T_\beta(\beta') \geq 0$ for all β' and $\sum_{\beta'} T_\beta(\beta') = 1-(\eta/2)$. The condition $T_\beta(\beta')\geq 0$ follows from observing that both the terms in the numerator of $T_\beta(\beta')$ are non-negative. The second condition follows from noting that

$$\sum_{\beta'} T_\beta(\beta') \;=\; \sum_{\beta'}\left(U_{C,\beta}(\beta') - \left(\frac{\eta}{2}\right)\cdot U_{k-1}(\beta')\right) \;=\; 1-\frac{\eta}{2}\,.$$

This gives a distribution $m_C^{(\beta)}$ on $\{-1,1\}^{S_C\setminus\{j\}}$ such that $U_{C,\beta} = (1-(\eta/2))\cdot m_C^{(\beta)} + (\eta/2)\cdot U_{k-1}$. ■

Thus, in the definition of the distributions m_S, the process of assigning the remaining $k-1$ variables in a constraint C conditioned on an assignment β to one of the variables, can be viewed as assigning them from the distribution $m_C^{(\beta)}$ with probability $1-(\eta/2)$ and from U_{k-1} with probability $\eta/2$. We can equivalently view the definition of the distribution m_S as first making the choice for every $C \in \mathcal{C}_G \cap B^{(d)}(S)$, whether conditioned on the parent of C, the rest of the variables in C will be assigned according to $m_C^{(\beta)}$ (which happens with probability $1-(\eta/2)$) or according to U_{k-1} (which happens with probability $\eta/2$).

For $S \subseteq \mathcal{V}_G$, we say an edge from a constraint $C \in \mathcal{C}_G \cap B^{(d)}(S)$ to a variable $x_{j'} \in \mathcal{V}_G \cap B^{(d)}(S)$ is *broken* in m_S, if $x_{j'}$ is assigned according to U_{k-1}, conditioned on some other variable x_j which is the parent of C in $B^{(d)}(S)$. Note that conditioned on the event that the edge from C to $x_{j'}$ is broken, the distributions of x_j and $x_{j'}$ are independent. This is because for any assignment β to x_j, $x_{j'}$ is assigned uniformly in $\{-1,1\}$.

We say that a path between two variables $x_j, x_{j'} \in \mathcal{V}_G \cap B^{(d)}(S)$ is broken if some edge in the path is broken. Note that if the length of the path is ℓ, then there $\ell/2$ edges going from a constraint vertex

to a variable vertex, and hence the path is broken with probability at least $1-(1-\eta/2)^{\ell/2}$. Also, we have as before that conditioned on the path between x_j and $x_{j'}$ being broken, the distributions of x_j and $x_{j'}$ are independent.

Recall that we are considering the case when $B^{(d)}(S) = \cup_{i=1}^{t} B^{(d)}(S_i)$ is a forest but $B^{(d)}(T) = B^{(d)}(S \cup \{v\})$ is a tree. Note that even though $v \notin S$, we can still have $v \in B^{(d)}(S)$. We first present the argument for the case when this does not happen.

- **Case 2a: $\mathbf{v \notin B^{(d)}(S)}$.** Since $v \notin B^{(d)}(S)$, any path from v to S_i for $i \in [t]$ must have length at least d. To analyze the distribution in this case, we first assume by Lemma 4.12 that the starting vertex for defining the distribution is v. We define the following event for the distribution m_T

$$\mathcal{E} \;:=\; \Big\{\forall i \in [t],\ \text{all paths in } B^{(d)}(T) \text{ from } v \text{ to } S_i \text{ are broken}\Big\}\,.$$

Since $B^{(d)}(T)$ is a tree, there is exactly one path from v to a node in S_i. The probability that the path is not broken is at most $(1-\eta/2)^{d/2}$. Hence,

$$\mathbb{P}\left[\overline{\mathcal{E}}\right] \;\le\; |S| \cdot (1-\eta/2)^{d/2} \;\le\; 2^{-\Omega(\eta d)}\,,$$

for $|S| = O(\eta d)$. Also, if $m_T(\alpha \mid \mathcal{E})$ denotes the probability of the assignment $\alpha \in \{-1,1\}^S$ given than the event $\mathcal{E}$ happens, then we can write

$$\begin{aligned} m_T(\alpha) \;&=\; \mathbb{P}\left[\mathcal{E}\right] \cdot m_T(\alpha \mid \mathcal{E}) \;+\; \mathbb{P}\left[\overline{\mathcal{E}}\right] \cdot m_T(\alpha \mid \overline{\mathcal{E}}) \\ &=\; m_T(\alpha \mid \mathcal{E}) \;\pm\; 2^{-\Omega(\eta d)}\,, \end{aligned}$$

where we use $a = b \pm c$ to denote $|a-b| \le c$. Since the distribution of vertices separated by a broken path is independent, conditioned on the event $\mathcal{E}$, the distribution for the sets $S_1, \ldots, S_r$ must be independent. Let α_i denote the restriction of the assignment α to the set S_i. We then have by the independence that

$$m_T(\alpha \mid \mathcal{E}) \;=\; \prod_{i=1}^{t} \left(m_T(\alpha_i \mid \mathcal{E})\right)\,.$$

Conditioned on the event $\mathcal{E}$, the assignment for the set S_i in the distribution m_T is defined by considering a subtree of $B^{(d)}(T)$ which does not include any vertices from S_j for $j \neq i$. The distribution on this subtree will be identical if we instead define the assignment according to the distribution $m_{S_i \cup \{v\}}$ conditioned on all paths from v to S_i in $B^{(d)}(S_i \cup \{v\})$ being broken. Let $\mathcal{E}_i$ denote this event. Then, since $\mathbb{P}\left[\mathcal{E}_i\right] \ge 1 - 2^{-\Omega(\eta d)}$, we have,

$$m_T(\alpha_i \mid \mathcal{E}) \;=\; m_{S_i \cup \{v\}}(\alpha_i \mid \mathcal{E}_i) \;=\; m_{S_i \cup \{v\}}(\alpha_i) \;\pm\; 2^{-\Omega(\eta d)}\,.$$

However, $B^{(d)}(S_i)$ is a tree and $B^{(d)}(S_i \cup \{v\})$ is also a tree and hence by Case 1 we have $m_{S_i \cup \{v\}}(\alpha_i) = m_{S_i}(\alpha_i)$. Combining the above and using the fact that the components $B^{(d)}(S_i)$ are disconnected in $B^{(d)}(S)$, we get

$$m_T(\alpha) \;=\; \prod_{i=1}^{t} \left(m_{S_i}(\alpha_i)\right) \;\pm\; 2^{-\Omega(\eta d)} \;=\; m_S(\alpha) \;\pm\; 2^{-\Omega(\eta d)}\,.$$

- **Case 2b: $\mathbf{v \in B^{(d)}(S)}$.** Without loss of generality, let $v \in B^{(d)}(S_1)$. The treatment for this case is almost identical except that we need to treat the set S_1 more carefully since the paths from v to S_1 may now be short. We now define the event $\mathcal{E}$ as

$$\mathcal{E} \;:=\; \left\{\forall i \in \{2,\ldots,t\},\ \text{all paths in } B^{(d)}(T) \text{ from } v \text{ to } S_i \text{ are broken}\right\}.$$

As before, conditioned on the event $\mathcal{E}$, the distributions on different sets S_i are independent and we can write

$$m_T(\alpha \mid \mathcal{E}) \;=\; m_T(\alpha \mid \mathcal{E}) \;\pm\; 2^{-\Omega(\eta d)} \;=\; \prod_{i=1}^{t} (m_T(\alpha_i \mid \mathcal{E})) \;\pm\; 2^{-\Omega(\eta d)}.$$

Again, we have that conditioned on the event $\mathcal{E}$, the distribution for the assignment α_i is defined by considering a subtree not containing any set S_j for $i \neq j$. For $i \geq 2$, letting $\mathcal{E}_i$ denote the event that all paths between v and S_i in $B^{(d)}(S_i \cup \{v\})$ are broken, we have

$$\prod_{i=1}^{t} (m_T(\alpha_i \mid \mathcal{E})) \;=\; m_{S_1\cup\{v\}}(\alpha_1) \cdot \prod_{i=2}^{t} \left(m_{S_i\cup\{v\}}(\alpha_i \mid \mathcal{E}_i)\right).$$

As before $m_{S_i\cup\{v\}}(\alpha_i \mid \mathcal{E}_i) = m_{S_i\cup\{v\}}(\alpha_i) \pm 2^{-\Omega(\eta d)}$ for all $i \geq 2$ and $m_{S_i\cup\{v\}}(\alpha_i) = m_{S_i}(\alpha_i)$ for all $i \in [t]$ by Case 1. Combining the above, we again get

$$m_T(\alpha) \;=\; \prod_{i=1}^{t} (m_{S_i}(\alpha_i)) \;\pm\; 2^{-\Omega(\eta d)} \;=\; m_S(\alpha) \;\pm\; 2^{-\Omega(\eta d)}.$$

■

Now that we have a family $\{m_S\}_{|S|\leq r}$ of approximately locally consistent probability distributions, we can use it to define locally consistent distribution $\{m'_S\}_{|S|\leq t}$ using a result by Raghavendra and Steurer [31].

Lemma 4.15 ([31]) *Let $\left\{m_S : \{-1,1\}^S \to \mathbb{R}_+\right\}_{|S|\leq r}$ be a family of probability distributions such that for all $S \subseteq T$ and $\alpha \in \{-1,1\}^S$:*

$$|m_S(\alpha) - m_T(\alpha)| \;\leq\; \varepsilon_0.$$

Then there exists a family of probability distributions $\left\{m'_S : \{-1,1\}^S \to \mathbb{R}_+\right\}_{|S|\leq r}$ such that for all $S \subseteq T$ and $\alpha \in \{-1,1\}^S$:

$$m'_S(\alpha) \;=\; m'_T(\alpha),$$

and for all S with $|S| \leq r$, we have $\|m_S - m'_S\|_1 \leq O(2^r \cdot \varepsilon_0)$.

Therefore, using Lemmas 4.13, 4.15, and 4.11, we can now prove the following theorem, which also completes our proof of Theorem 2.17.

Theorem 4.16 *Let $f : \{-1,1\}^k \to \{0,1\}$ be a predicate such that there exists a vanishing measure Λ^* on $\mathcal{C}^*(f)$. Then, for every $\varepsilon > 0$, there is a constant $c_\varepsilon > 0$, such that for all large enough n, there exists an instance of $\mathsf{CSP}(f)$ on n variables satisfying the following:*

- *For any integral assignment ψ, the fraction of the constraints satisfied is in the range $[\rho(f) - \varepsilon, \rho(f) + \varepsilon]$.*

- *The optimum for the linear program obtained by* $c_\varepsilon \cdot \log\log n$ *rounds of the Sherali-Adams hierarchy is at least* $1 - O(k \cdot \sqrt{\varepsilon})$.

Proof: The proof follows simply from appropriate choices for the parameters η, d, r and δ. Using Lemma 4.6 we obtain an instance such that the constraint graph G has girth $g = O(\log n)$, degree $D = O_\varepsilon(1)$ and such that the fraction of constraints satisfied by any integral assignment ψ is between $\rho(f) - \varepsilon$ and $\rho(f) + \varepsilon$.

Using Claim 4.10, we can define the distributions m_S for all sets of size at most r, when $r \cdot D^d \leq g$. Setting $\delta = \sqrt{\varepsilon}$ and $\eta = k \cdot \sqrt{\varepsilon}$, we get from Lemma 4.11 that the distributions m_S achieve LP value $1 - O(k\sqrt{\varepsilon})$. Taking the error $2^{-\Omega(\eta d)}$ from Lemma 4.13 to be ε_0 in Lemma 4.15, and r to be such that $2^r \cdot \varepsilon_0 = O(\sqrt{\varepsilon})$, we get that the LP value achieved by the the distributions m'_S is at least $O(1 - k\sqrt{\varepsilon})$. Since we only need that $r \cdot D^d \leq g$ for defining the distributions and $2^r \cdot 2^{-\Omega(\eta d)} = O(\sqrt{\varepsilon})$ for bounding the error, we can choose both d and r to be $\Omega_\varepsilon(\log g) = \Omega_\varepsilon(\log\log n)$. ■

Acknowledgements

The first author would like to thank Per Austrin and Johan Håstad for many discussions over the years on the topic of approximation resistance.

References

[1] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof Verification and Hardness of Approximation Problems. In *FOCS*, pages 14–23, 1992.

[2] Sanjeev Arora and Shmuel Safra. Probabilistic Checking of Proofs; A New Characterization of NP. In *FOCS*, pages 2–13, 1992.

[3] Per Austrin and Johan Håstad. Randomly Supported Independence and Resistance. *SIAM J. Comput.*, 40(1):1–27, 2011.

[4] Per Austrin and Johan Håstad. On the Usefulness of Predicates. In *IEEE Conference on Computational Complexity*, pages 53–63, 2012.

[5] Per Austrin and Subhash Khot. A characterization of approximation resistance for even k-partite csps. In *ITCS*, pages 187–196, 2013.

[6] Per Austrin and Elchanan Mossel. Approximation Resistant Predicates from Pairwise Independence. *Computational Complexity*, 18:249–271, 2009.

[7] Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, pcps, and nonapproximability-towards tight results. *SIAM J. Comput.*, 27(3):804–915, 1998.

[8] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP Gaps from Pairwise Independence. *Theory of Computing*, 8(1):269–289, 2012.

[9] Siu On Chan. Approximation Resistance from Pairwise Independent Subgroups. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:110, 2012.

[10] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Integrality Gaps for Sherali-Adams Relaxations. In *STOC*, pages 283–292, 2009.

[11] Mahdi Cheraghchi, Johan Håstad, Marcus Isaksson, and Ola Svensson. Approximating Linear Threshold Predicates. *TOCT*, 4(1):2, 2012.

[12] Eden Chlamatac and Madhur Tulsiani. Convex Relaxations and Integrality Gaps. In *Handbook on Semidefinite, Cone and Polynomial Optimization*, 2010.

[13] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear Programming Relaxations of Max-Cut. In *SODA*, pages 53–61, 2007.

[14] Lars Engebretsen, Jonas Holmerin, and Alexander Russell. Inapproximability Results for Equations over Finite Groups. *Theor. Comput. Sci.*, 312(1):17–45, 2004.

[15] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Approximating Clique is Almost NP-Complete (Preliminary Version). In *FOCS*, pages 2–12, 1991.

[16] Uriel Feige and Gideon Schechtman. On the Optimality of the Random Hyperplane Rounding Technique for MAX CUT. *Random Struct. Algorithms*, 20(3):403–440, 2002.

[17] Michel X. Goemans and David P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *J. ACM*, 42(6):1115–1145, 1995.

[18] Venkatesan Guruswami, Daniel Lewin, Madhu Sudan, and Luca Trevisan. A tight characterization of np with 3 query pcps. In *FOCS*, pages 8–17, 1998.

[19] Gustav Hast. *Beating a Random Assignment*. PhD thesis, Royal Institute of Technology, Sweden, 2005.

[20] Gustav Hast. Beating a Random Assignment. In *APPROX-RANDOM*, volume 3624 of *Lecture Notes in Computer Science*, pages 134–145, 2005.

[21] Johan Håstad. Every 2-CSP Allows Nontrivial Approximation. *Computational Complexity*, 17(4):549–566, 2008.

[22] Steve Horvath. *Weighted Network Analysis*. Springer-Verlag, 2011.

[23] Subhash Khot. Hardness Results for Coloring 3 -Colorable 3 -Uniform Hypergraphs. In *FOCS*, pages 23–32, 2002.

[24] Subhash Khot. On the Power of Unique 2-Prover 1-Round Games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, STOC '02, pages 767–775, New York, NY, USA, 2002. ACM.

[25] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM J. Comput.*, 37(1):319–357, 2007.

[26] Subhash Khot and Rishi Saket. SDP Integrality Gaps with Local ℓ_1-Embeddability. In *FOCS*, pages 565–574, 2009.

[27] Mikhail Lifshits. *Lectures on Gaussian Processes*. Springer-Verlag, Springer Briefs in Mathematics, 2012.

[28] Ryan O'Donnell and Yi Wu. An optimal sdp algorithm for max-cut, and equally optimal long code tests. In *STOC*, pages 335–344, 2008.

[29] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *STOC*, pages 245–254, 2008.

[30] Prasad Raghavendra and David Steurer. How to Round Any CSP. In *FOCS*, pages 586–594, 2009.

[31] Prasad Raghavendra and David Steurer. Integrality Gaps for Strong SDP Relaxations of UNIQUE GAMES. In *FOCS*, pages 575–585, 2009.

[32] Ran Raz. A Parallel Repetition Theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[33] Alex Samorodnitsky and Luca Trevisan. A PCP Characterization of NP with Optimal Amortized Query Complexity. In *STOC*, pages 191–199, 2000.

[34] Thomas J. Schaefer. The complexity of Satisfiability Problems. In *STOC*, pages 216–226, 1978.

[35] Grant Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k-CSPs. In *FOCS*, pages 593–602, 2008.

[36] Johan Håstad. On the Efficient Approximability of Constraint Satisfaction Problems. In *Surveys in Combinatorics*, volume 346, pages 201–222. Cambridge University Press, 2007.

[37] Michael Taylor. *Measure Theory and Integration*. AMS, Graduate Studies in Mathematics, 2006.

[38] Uri Zwick. Approximation Algorithms for Constraint Satisfaction Problems Involving at Most Three Variables per Constraint. In *SODA*, pages 201–210, 1998.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de