Randomly Supported Independence and Resistance

Per Austrin*
KTH – Royal Institute of Technology
Stockholm, Sweden
austrin@kth.se

Johan Håstad
KTH – Royal Institute of Technology
Stockholm, Sweden
johanh@kth.se

ABSTRACT

We prove that for any positive integer k, there is a constant c_k such that a randomly selected set of $c_k n^k \log n$ Boolean vectors with high probability supports a balanced k-wise independent distribution. In the case of $k \leq 2$ a more elaborate argument gives the stronger bound $c_k n^k$. Using a recent result by Austrin and Mossel this shows that a predicate on t bits, chosen at random among predicates accepting $c_2 t^2$ input vectors, is, assuming the Unique Games Conjecture, likely to be approximation resistant.

These results are close to tight: we show that there are other constants, c'_k, such that a randomly selected set of cardinality $c'_k n^k$ points is unlikely to support a balanced k-wise independent distribution and, for some $c > 0$, a random predicate accepting $ct^2/\log t$ input vectors is non-trivially approximable with high probability.

In a different application of the result of Austrin and Mossel we prove that, again assuming the Unique Games Conjecture, *any* predicate on t bits accepting at least $(32/33) \cdot 2^t$ inputs is approximation resistant.

The results extend from the Boolean domain to larger finite domains.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]

General Terms

Theory

1. INTRODUCTION

The motivation of this paper comes from the approximability of maximum constraint satisfaction problems (Max-CSPs). A problem is defined by a t-ary predicate P and an instance is given by a list of t-tuples of literals over Boolean variables[1]. The task is to find an assignment to the variables such that as many as possible of the t-tuples of literals satisfy the predicate P.

The most famous such problem is probably Max-3-Sat where $t = 3$ and P is simply the disjunction of the three bits. Another problem that (almost) falls into this category is Max-Cut, in which $t = 2$ and P is non-equality. In traditional Max-Cut we do not allow negated literals and if we do allow negation the problem becomes Max-2-Lin-2, linear equations modulo 2 with two variables in each equation.

These two problems, as well as almost all Max-CSPs, are NP-hard and the main focus of research on these problems has been approximation algorithms. An algorithm is considered to be a C-approximation if it, on each input, finds an assignment with an objective value that is within a factor C of the optimal solution. We allow randomized algorithms and in this case it is sufficient that the expected value of the objective values satisfies the desired bound.

To define what is non-trivial is a matter of taste but hopefully there is some consensus that the following algorithm is trivial: Without looking at the instance pick a random value for each variable. We say that an approximation ratio C is non-trivial if it is better than the ratio obtained by this trivial algorithm. We call a predicate *approximation resistant* if it is NP-hard to achieve a non-trivial approximation ratio.

It is perhaps surprising but many CSPs are approximation resistant and one basic example is Max-3-Sat [12]. The famous approximation algorithm of Goemans and Williamson [9] shows that Max-Cut is not approximation resistant and this result can be extended in great generality to show that no predicate that depends on two inputs from an arbitrary finite domain can be approximation resistant [13].

Zwick [22] established approximability results for predicates that depend on three Boolean inputs and from this it follows that the only predicates on three inputs that are approximation resistant are those that are implied by parity or its negation. Many scattered results on wider predicates do exist [10, 19] and in particular Hast [11] made an extensive classification of predicates on four inputs.

These results for predicates of small width give little guidance on what to expect for a generic predicate. Generally speaking there are several results pointing towards the direction that predicates that accept more inputs are more likely to be approximation resistant. We say that a predicate P implies a predicate Q if any assignment that satisfies P also satisfies Q. We say that a predicate P is *heredi-*

*Work done in part while the author was visiting U.C. Berkeley under a grant from the Swedish Royal Academy of Sciences.

STOC'09, May 31–June 2, 2009, Bethesda, Maryland, USA.

[1]Our results extend to larger domains but for simplicity in the extended abstract we stay with the Boolean domain.

tarily approximation resistant if any predicate implied by P is approximation resistant. Most predicates known to be approximation resistant also turn out to be hereditarily approximation resistant. One of the few predicates that does not have this property is $P(x_1, x_2, x_3, x_4)$ which is satisfied if x_1 is true and $x_2 \neq x_3$ or x_1 is false and $x_2 \neq x_4$. This was proved approximation resistant by Guruswami et al. [10] but implies $NAE(x_2, x_3, x_4)$ which admits a nontrivial approximation algorithm, see for instance [22].

As a generic positive result Hast [11] proved that any predicate on t bits that accepts fewer than $2\lceil (t+1)/2 \rceil$ inputs does admit a nontrivial approximation algorithm. This might at first seem like a rather weak result but evidence is mounting that this is very close to the best possible result of this type. Let us elaborate on this evidence.

The strongest inapproximability results depend on the Unique Games Conjecture, UGC, of Khot [16]. The truth of this conjecture is still very much open and probably the most important open problem in the theory of approximability. Even if we should not take a hardness result based on UGC as a final word it is a very valuable result. Despite many strong efforts to disprove the conjecture [21, 6, 1], the conjecture remains open. As these results appear to push the currently available algorithmic techniques as far as they can go, any negative result based on the UGC rules out an algorithm using current techniques and thus it is a strong indication that a problem is difficult.

Using the UGC, Samorodnitsky and Trevisan [20] proved that when t is of the form $2^r - 1$, Hast's result is tight and there is an approximation resistant predicate that accepts $t+1$ inputs. The proof extends to give hereditary approximation resistance and using this Håstad [14] proved that a predicate chosen at random from all predicates that accept s inputs is likely to be approximation resistant if $s = \omega(2^t/\sqrt{t})$. For $t = 2^r - 1$ the bound on s can be as low as $2^t/t$ but this is the lower limit of what can be obtained using the predicates of Samorodnitsky and Trevisan.

Austrin and Mossel [3], using the machinery of Mossel [18] extended the results of Samorodnitsky and Trevisan to apply to a much wider class of predicates. To be more precise they proved that any predicate P for which there exists a balanced pairwise independent distribution supported on the inputs accepted by P is, assuming the UGC, hereditarily approximation resistant. Using this they proved that without assumptions on the form of t there are predicates that accept $t + o(t)$ inputs which satisfy this property. Furthermore if the famous conjecture on the existence of Hadamard matrices is true their bound is $4\lceil (t+1)/4 \rceil$, matching the bounds of Hast for half of all values of t and being off by an additive constant of 2 for other values.

The result of Austrin and Mossel is very powerful and we use it as a tool to investigate the approximation resistance of randomly chosen predicates. The technical question that arises is to analyze the probability that s random Boolean vectors of length t can support a balanced pairwise independent distribution, and in particular for what values of s this probability is $1 - o(1)$. Many properties of pairwise independent distributions have been studied, but we have not found any results on randomly supported pairwise independent distributions. We feel that this is natural question interesting in its own right and we study the question in some generality, looking at the question of existence of a k-wise independent distribution establishing the following result.

THEOREM 1.1. *(informal) There are absolute constants c_k such that if we pick $c_k n^k \log n$ random Boolean vectors of length n, then with high probability there is a k-wise independent distribution supported on these points.*

For the case $k = 2$, which is most important for our application, we are able to remove the logarithmic factor, obtaining the following result.

THEOREM 1.2. *(informal) There is an absolute constant c_2 such that if we pick $c_2 n^2$ random Boolean vectors of length n, then with high probability there is a pairwise independent distribution supported on these points.*

We remark that for the case of supporting an unbiased probability distribution, i.e., the case $k = 1$, a sharp bound of $2n$ on the threshold is already known by an elegant result by Füredi [8].

The bounds for the case $k \leq 2$ are asymptotically tight: we prove that for any constant k, $\Omega(n^k)$ random strings are needed to have a good probability to be the support of a k-wise independent probability distribution.

Through the result of Austrin and Mossel the existence of a pairwise independent distribution gives approximation resistance and we have the following immediate corollary.

COROLLARY 1.3. *(informal) There is an absolute constant c_2 such that if we pick a random predicate P on t bits which accepts $c_2 t^2$ of the 2^t possible input strings then, assuming the UGC, with high probability P is hereditarily approximation resistant.*

Even though we have a tight answer for the number of points needed to support a pairwise independent distribution this does automatically give an answer to the question when a predicate is approximation resistant. Here we get an almost tight result by showing that, for some constant $c > 0$, a predicate that accepts a random set of size $ct^2/\log t$ is likely to admit a nontrivial approximation algorithm.

This result follows by an application of an algorithm of Hast [11]. Broadly speaking the algorithm looks at the "quadratic part" of the predicate and applies a standard semidefinite programming approach.

All these results have looked at very sparse sets. For rather dense sets we can prove similar results with certainty.

THEOREM 1.4. *There is a constant c such that any subset of size $(1 - c^k)2^n$ of the Boolean cube supports a k-wise independent distribution.*

For the case of $k = 2$ we are interested in an explicit value of the constant and we have the following corollary.

COROLLARY 1.5. *Any predicate on t bits that accepts more than $(32/33) \cdot 2^t$ inputs is, assuming the UGC, approximation resistant.*

The best previous results of this form are that any predicate accepting more than $2^t(1 - 2^{-\sqrt{t}})$ inputs is resistant assuming $P \neq NP$ [11], and that any predicate accepting more than $2^t(1 - (2t)^{-1})$ inputs is resistant assuming the UGC [14].

The constant 32/33 in Corollary 1.5 is not tight. A lower bound on the correct value of this constant is 13/16: Hast [11] gives a non-trivially approximable predicate on 4 variables

which accepts 13 of the 16 assignments. For the corresponding constant in Theorem 1.4 for $k = 2$, the correct value is strictly larger than 13/16 (we elaborate on this in the full version of the paper).

The results extend to arbitrary finite domains, and to non-balanced k-wise independence and most proofs for the general case can be found in [2].

An outline of the paper is as follows. After giving preliminaries in Section 2 and Section 3 we establish Theorem 1.4 and Corollary 1.5 in Section 4. In Section 5 we prove the upper bound on the size of random support for a k-wise independent distribution and give the stronger bound for pairwise independence in Section 6. For the reverse directions we give the lower bound on the number of random points needed to support a k-wise independent distribution in Section 7 and the approximation result for sparse predicates in Section 8. We end with some conclusions in Section 9 and some standard proofs appears in an appendix.

2. PRELIMINARIES

We use $\{-1,1\}^n$ to denote the n-dimensional Boolean hypercube, or equivalently, the set of binary strings of length n (as is common, we use ± 1 to represent bits rather than 0, 1, as this simplifies the computations). We denote by $\mathcal{U}_n$ the uniform distribution over $\{-1,1\}^n$.

For a function $f : \{-1,1\}^n \to \mathbb{R}$, and $1 \leq p < \infty$, we denote its ℓ_p *norm* by $||f||_p = (\mathbb{E}_x[|f(x)|^p])^{1/p}$, where the expected value is with respect to the uniform distribution on $\{-1,1\}^n$. The ℓ_∞ norm of f is defined by $||f||_\infty = \max_x |f(x)|$. We remind the reader of *Hölder's Inequality*: let $1 \leq p, q \leq \infty$ be such that $1/p + 1/q = 1$, and let $f, g : \{-1,1\}^n \to \mathbb{R}$. Then

$$\mathop{\mathbb{E}}_x[f(x)g(x)] \leq ||f||_p \cdot ||g||_q$$

For a probability distribution μ over $\{-1,1\}^n$ and subset $S \subseteq [n]$ of coordinates, we denote by μ_S the *marginal distribution* of μ on the coordinates in S (i.e., the distribution on $\{-1,1\}^{|S|}$ induced by μ by only looking at the coordinates in S). A distribution μ over $\{-1,1\}^n$ is *balanced k-wise independent* if, for every $S \subseteq [n]$ with $|S| = k$, it holds that $\mu_S = \mathcal{U}_k$. For a probability distribution μ over $\{-1,1\}^n$, we denote by $\mathrm{Supp}(\mu) = \{\, x : \mu(x) > 0 \,\}$ the *support* of μ.

For vectors $u, v \in \mathbb{R}^n$, we denote by $\langle u, v\rangle = \sum_{i=1}^n u_i v_i$ their *inner product*. We denote by $\mathbf{0} = \mathbf{0}_n \in \mathbb{R}^n$ the all-zeros vector in $\mathbb{R}^n$, and always drop the subscript n as the dimension will be clear from the context.

Given a set $X \subseteq \mathbb{R}^n$, $\mathrm{Conv}(X)$ denotes the *convex hull* of X, defined as the smallest convex set containing X. For $X = \{x_1, \ldots, x_m\}$ finite, $\mathrm{Conv}(X)$ is the set of all points which are convex combinations of $x_1, \ldots, x_m$,

$$\mathrm{Conv}(X) = \left\{ \sum_{i=1}^m \alpha_i x_i \,:\, \alpha_i \geq 0, \sum_{i=1}^m \alpha_i = 1 \right\}.$$

We will also need the following standard result on small ϵ-nets of the unit sphere (see e.g. [17]):

THEOREM 2.1. *For every n and $0 < \epsilon < 1/3$, there exists a set S of at most $(5/\epsilon)^n$ unit vectors in $\mathbb{R}^n$, such that, for any unit vector $u \in \mathbb{R}^n$, there is a $v \in S$ satisfying*

$$\langle u, v\rangle \geq 1 - \epsilon.$$

2.1 Multilinear Polynomials

We shall frequently work with multilinear polynomials $f : \{-1,1\}^n \to \mathbb{R}$, defined in terms of

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i,$$

for some set of coefficients $\{\hat{f}_S\}_{S \subseteq [n]}$. It is well-known that any function $f : \{-1,1\}^n \to \mathbb{R}$ can be uniquely represented as such a multilinear polynomial, and $\hat{f}(S)$ are the Fourier-Walsh coefficients of f. Henceforth we shall refer to functions $f : \{-1,1\}^n \to \mathbb{R}$ as polynomials. We also introduce the convenient shorthand

$$\chi_S(x) = \prod_{i \in S} x_i$$

for the multilinear monomial corresponding to the set $S \subseteq [n]$.

We say that a polynomial $f : \{-1,1\}^n \to \mathbb{R}$ has degree d if $\hat{f}(S) = 0$ for every S with $|S| > d$. We let $f^{=d}$ denote the part of f that is of degree exactly d.

As we frequently work with polynomials f of low degree, say k, and constant coefficient $\hat{f}(\emptyset) = 0$, we introduce the following notation for the set of all $S \subseteq [n]$ with cardinality $1 \leq |S| \leq k$:

$$D_k := D_k(n) = \{\, S \subseteq [n] \,|\, 1 \leq |S| \leq k \,\},$$

and denote by $d_k := d_k(n)$ the cardinality $d_k = |D_k|$. Note that $d_k = \sum_{i=1}^k \binom{n}{i} \leq n^k$.

It is useful to view the monomials that can be input into a low degree polynomial as a vector and towards this end let us introduce the following notation.

DEFINITION 2.2. *Given a string $x \in \{-1,1\}^n$, we define $x^{:\leq k:}$ as*

$$x^{:\leq k:} = \bigoplus_{S \in D_k} \chi_S(x) \in \{-1,1\}^{d_k},$$

Here, $\oplus$ denotes the direct sum, e.g., $a \oplus b \oplus c = (a, b, c)$. In other words, $x^{:\leq k:}$ is the vector obtained by writing down the values of all non-constant monomials of degree at most k, evaluated at x. For a set $X \subseteq \{-1,1\}^n$, we use $X^{:\leq k:} \subseteq \{-1,1\}^{d_k} \subseteq \mathbb{R}^{d_k}$ to denote the set $\{\, x^{:\leq k:} \,|\, x \in X \,\}$.

Note that every $v \in \mathbb{R}^{d_k}$ is in 1–1 correspondence with a degree-k polynomial $f_v : \{-1,1\}^n \to \mathbb{R}$ with $\mathbb{E}[f_v] = 0$, defined by $f_v(x) = \langle v, x^{:\leq k:}\rangle$ for every $x \in \{-1,1\}^n$ (i.e., we interpret v as the Fourier-Walsh coefficients of f_v).

2.2 Hypercontractivity

The main analytic tool in all our upper bounds is *hypercontractivity*. A well-known consequence of the famous *Hypercontractivity Theorem* [5, 4] can be stated as follows.

THEOREM 2.3. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial. Then, for every $1 \leq p < q \leq \infty$, it holds that*

$$||f||_p \geq \sqrt{\frac{p-1}{q-1}}^d ||f||_q.$$

The following stronger estimate for the case $p = 2$, $q = 4$, and $d = 2$ (i.e., quadratic polynomials) is sometimes useful.

THEOREM 2.4. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a degree-2 polynomial. Then*

$$||f||_2 \geq 15^{-1/4} ||f||_4.$$

This estimate is not new, but as we do not know of a reference for it, we include a proof. A (different) proof of the same inequality for degree-2 multilinear polynomials in Gaussian variables can be found in [15], Corollary 7.36 and Remark 7.37.

PROOF. We want to estimate $\mathbb{E}[f^4]$ for a quadratic polynomial f. We do this by expanding the fourth power and looking at the expectation of each term. Any term that contains a variable to an odd power gives zero contribution to the expected value and thus we only care about terms of even degree. Replacing any linear terms x_i by $x_0 x_i$ for a new variable x_0 we get the same expected value and hence we can assume that f is homogeneous of degree two. For notation let us use $f(x) = \sum_e \hat{f}_e x_i x_j$ for edges $e = (i,j)$ and let us order the edges in the lexicographic order.

Let us look at the expansion of f^4. We have the following three types of terms that contribute to the expected value:

1. $\hat{f}_e^4$.
2. $\hat{f}_{e_1}^2 \hat{f}_{e_2}^2$ with $e_1 < e_2$.
3. $\hat{f}_{e_1}\hat{f}_{e_2}\hat{f}_{e_3}\hat{f}_{e_4}$ with all edges e_i distinct and forming a quadrilateral.

The first type of terms appear with coefficient 1, the second type with coefficient 6 and the last with coefficient 24.

Let us apply the inequality $ab \leq \frac{1}{2}(a^2+b^2)$ for the terms of type three with a the product of two edges without common endpoints. This gives new terms of the form $\hat{f}_{e_1}^2 \hat{f}_{e_2}^2$. Given e_1 and e_2 there are two ways to choose (e_3, e_4) to complete the quadrilateral. Both of these choices gives a contribution $12\hat{f}_{e_1}^2 \hat{f}_{e_2}^2$ and thus we get the total estimate

$$\sum_e \hat{f}_e^4 + 30 \sum_{e_1<e_2} \hat{f}_{e_1}^2 \hat{f}_{e_2}^2,$$

for $\mathbb{E}[f^4]$. This is clearly bounded by $15(\sum_e \hat{f}_e^2)^2 = 15\,\mathbb{E}[f^2]^2$ and the proof is complete. □

For some of our proofs, we need that the ℓ_1 norm is related to the ℓ_2 norm, which is not an immediate consequence of Theorem 2.3. It does however follow from a classic "duality" argument.

THEOREM 2.5. *Let f be a random variable. If f satisfies $||f||_2 \geq \delta ||f||_p$ for some constants $p > 2$ and $\delta > 0$, then*

$$||f||_1 \geq \delta^{p/(p-2)} ||f||_2.$$

PROOF. Let $r = (p-2)/(2p-2) \in (0, 1/2)$, and define $g(x) = f(x)^{2r}$, $h(x) = f(x)^{2-2r}$. By Hölder's Inequality,

$$\begin{aligned} ||f||_2^2 &\leq ||g||_{1/2r} \cdot ||h||_{1/(1-2r)} = ||f||_1^{2r} \cdot ||f||_{(2-2r)/(1-2r)}^{2-2r} \\ &= ||f||_1^{2r} \cdot ||f||_p^{2-2r} \leq \delta^{2r-2} \cdot ||f||_1^{2r} \cdot ||f||_2^{2-2r} \end{aligned}$$

Simplifying, we get $||f||_1 \geq \delta^{(1-r)/r} ||f||_2 = \delta^{p/(p-2)} ||f||_2$. □

Combined with the Hypercontractivity Theorem, this gives

THEOREM 2.6. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial. Then*

$$||f||_1 \geq e^{-d} ||f||_2.$$

PROOF. The Hypercontractivity Theorem combined with Theorem 2.5 implies that for any $p > 2$,

$$||f||_1 \geq \left((p-1)^{-\frac{p}{2(p-2)}}\right)^d ||f||_2.$$

Letting $p \to 2$, this gives the desired bound. □

2.3 Concentration Bounds

It is known that hypercontractivity implies good concentration bounds for low-degree polynomials (see e.g. [7]). We will need the following two results, the standard proofs of which can be found in the appendix.

THEOREM 2.7. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial with $||f||_2 = 1$ and any $t > e^{d/2}$,*

$$\Pr[|f| > t] \leq \exp(-ct^{2/d}),$$

where $c := \frac{d}{2e}$.

THEOREM 2.8. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a degree-2 polynomial with $||f||_2 = 1$, and let $x_1, \ldots, x_m$ be a sequence of m independent uniformly random elements of $\{-1,1\}^n$. Then, for every $r > 0$ satisfying $r < 2e\sqrt{m}$, it holds that*

$$\Pr\left[\left|\sum_{i=1}^m f(x_i) - m\,\mathbb{E}[f]\right| > r\sqrt{m}\right] \leq 2\exp\left(-\frac{r^2}{8e^2}\right).$$

Furthermore, this holds also if f is replaced by $|f|$.

3. LIMITED INDEPENDENCE AND LOW-DEGREE POLYNOMIALS

We now characterize the sets $X \subseteq \{-1,1\}^n$ which support k-wise independent distributions, in terms of degree-k polynomials over $\{-1,1\}^n$.

THEOREM 3.1. *Let $X \subseteq \{-1,1\}^n$ be a set of binary strings. Then, the following conditions are equivalent:*

(1) There exists a balanced k-wise independent distribution μ over $\{-1,1\}^n$ such that $\mathrm{Supp}(\mu) \subseteq X$.

(2) $\mathbf{0} \in \mathrm{Conv}(X^{:\leq k:})$.

(3) There is no degree k polynomial $f : \{-1,1\}^n \to \mathbb{R}$ such that $f(x) > \mathbb{E}[f]$ for every $x \in X$.

This characterization is most likely already known, but as we have not been able to find it in the literature, we give a proof here.

PROOF. **(1) ⇔ (2).** We view $\mathrm{Conv}(X^{:\leq k:})$ as the set of probability distributions over $\{-1,1\}^n$ supported on X. Any convex combination $\sum_{x\in X} \alpha_x \cdot x^{:\leq k:} \in \mathrm{Conv}(X^{:\leq k:})$ corresponds to the probability distribution μ_α over $\{-1,1\}^n$ in which

$$\mu_\alpha(x) = \begin{cases} \alpha_x & \text{if } x \in X \\ 0 & \text{otherwise} \end{cases}.$$

Thus, it suffices to prove that, for every convex combination $\{\alpha_x\}_{x\in X}$, the corresponding distribution μ_α has all k-dimensional marginals being the uniform distribution iff $\sum \alpha_x \cdot x^{:\leq k:} = \mathbf{0}$. This follows from the well known fact that a set of bits has the uniform distribution iff the exclusive-or of any subset is unbiased.

(2) ⇔ (3). Without loss of generality, we can restrict our attention to f such that $\mathbb{E}[f] = 0$. Now, $\mathbf{0}$ is *not* in the convex hull of $X^{:\leq k:}$ if and only if there exists a separating hyperplane $v \in \mathbb{R}^{d_k}$ such that $\langle v, x^{:\leq k:}\rangle > 0$ for every $x \in X$. The equivalence now follows by the correspondence between $v \in \mathbb{R}^{d_k}$ and degree-k polynomials f with $\mathbb{E}[f] = 0$. □

4. POLYNOMIALS ARE SOMEWHAT BALANCED

In this section we prove that low-degree polynomials must exceed their expectation by a constant amount on a constant fraction of inputs.

THEOREM 4.1. *There is a universal constant c such that for any degree-d polynomial $f : \{-1,1\}^n \to \mathbb{R}$ with $\mathbb{E}[f] = 0$ and $\mathrm{Var}[f] = 1$,*

$$\Pr[f > c^d] > c^d.$$

A similar statement can be found in [7]. They lower bound $\Pr[f > 0]$ rather than $\Pr[f > c^d]$, but this difference is superficial, and their proof (which is quite different from the one below) can be adapted to a proof of Theorem 4.1 as well.

PROOF. We are going to use the relation between the ℓ_1 norm and the ℓ_2 norm given by Theorem 2.6. Let $c = (4e)^{-2}$ and define $g : \{-1,1\}^n \to \mathbb{R}$ by

$$g(x) = \mathbf{1}_{f>c^d}(x) \cdot f(x) = \begin{cases} f(x) & \text{if } f(x) > c^d \\ 0 & \text{otherwise} \end{cases}.$$

We will lower bound $\Pr[f > c^d] = \Pr[g > 0]$ by the second moment method:

$$\Pr[g > 0] \geq \frac{\mathbb{E}[g]^2}{\mathbb{E}[g^2]} > ||g||_1^2, \tag{1}$$

where the last inequality follows from $\mathbb{E}[g^2] < \mathbb{E}[f^2] = 1$. For $||g||_1$, note that, since $\mathbb{E}[f] = 0$, we have $\mathbb{E}[\mathbf{1}_{f>0} \cdot f] = \frac{1}{2}||f||_1$, implying that

$$||g||_1 = \mathbb{E}[g] = \frac{1}{2}||f||_1 - \mathbb{E}[\mathbf{1}_{0<f\leq c^d} f] \geq \frac{1}{2}||f||_1 - c^d,$$

which, by Theorem 2.6, is lower-bounded by

$$||g||_1 \geq \frac{1}{2}e^{-d}||f||_2 - c^d \geq \frac{1}{2}e^{-d} - \frac{1}{4}e^{-2d} \geq \frac{1}{4}e^{-d} \geq c^{d/2}$$

so that $\Pr[g > 0] > ||g||_1^2 \geq c^d$, as desired. □

As an easy corollary, we see that for every k, any set $X \subseteq \{-1,1\}^n$ of sufficiently large constant density supports a k-wise independent distribution.

COROLLARY 4.2. *Every set $X \subseteq \{-1,1\}^n$ of size $|X| \geq 2^n(1 - e^{-2k}/4)$ supports a balanced k-wise independent distribution.*

The proof is a direct consequence of Theorem 3.1 and (the proof of) Theorem 4.1. As the corollary only needs a bound on $\Pr[f > 0]$ we define g to be the positive part of f. Then

$$||g||_1 = \frac{1}{2}||f||_1 \geq \frac{1}{2}e^{-k}||f||_2 = \frac{1}{2}e^{-k}$$

and the corollary follows from (1).

We note that the exponential dependence on the degree (i.e., the amount of independence) in both Theorem 4.1 and Corollary 4.2 is tight. To see this, consider a scaled version of the degree-d polynomial $f : \{-1,1\}^n \to \mathbb{R}$ defined by

$$f(x) = \prod_{i=1}^{d}(1 - x_i) - 1,$$

which takes the value $2^d - 1$ with probability 2^{-d}, and the value -1 with probability $1 - 2^{-d}$.

The bound in Corollary 4.2 is based on the relation between the ℓ_2 norm and the ℓ_1 norm. Using Theorems 2.4 and 2.5 one gets the bound $||f||_1 \geq 15^{-1/2}||f||_2$ for degree-2 polynomials. This in turn improves the bound for $k = 2$ in Corollary 4.2 from $1 - e^{-4}/4$ to $59/60$. As an alternative approach Ryan O'Donnell has suggested the following proof along the lines of the proof [7] for their variant of Theorem 4.1, giving an even better bound of $32/33$.

THEOREM 4.3. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a degree-2 polynomial with $\mathbb{E}[f] = 0$, $\mathrm{Var}[f] = 1$. Then $\Pr[f > 0] > 1/33$.*

PROOF. The proof is based on the inequality $\mathbf{1}_{x>0} \geq 0.13x + 0.062x^2 - 0.0021x^4$, where $\mathbf{1}_{x>0}$ is the indicator function of the event $x > 0$. Hence, we have that

$$\Pr[f(x) > 0] = \mathbb{E}[\mathbf{1}_{f(x)>0}] \geq 0.062\,\mathbb{E}[f^2] - 0.0021\,\mathbb{E}[f^4].$$

Using Theorem 2.4 to bound the ℓ_4 norm in terms of the ℓ_2 norm and plugging in $||f||_2 = 1$, we have that

$$\Pr[f(x) > 0] \geq 0.062 - 15 \cdot 0.0021 = 0.0305 > 1/33$$

We remark that choosing the coefficients more carefully, the lower bound of 0.0305 can be marginally improved (to roughly 0.0309401). □

Combining the proof above with the result of Austrin and Mossel [3] we get the following theorem.

THEOREM 4.4. *Let P be any predicate on t bits that accepts at least $(32/33) \cdot 2^t$ input strings. Then, assuming the UGC, P is approximation resistant.*

Theorem 4.3 uses the relation between ℓ_2 norm and ℓ_4 norm given by Theorem 2.4, and that bound is tight, so it is not clear whether the constant can be improved using this method. The first approach, giving $59/60$, uses the relation between ℓ_1 norm and ℓ_2 norm, for which our constant $15^{-1/2}$ is probably not the best possible. It is quite possible that that constant can be taken larger than $(33/4)^{-1/2}$, which would result in a better constant in Theorem 4.4.

5. OBTAINING K-WISE INDEPENDENCE

In this section, we give an upper bound of $(cn)^k \log(n^k)$ on the threshold for randomly supported independence. This comes relatively close to matching our lower bound of $\Omega(n^k)$ for constant k, being only a logarithmic factor off from being tight. In the next section, we prove our main theorem, that in the case $k = 2$, this logarithmic factor can be removed.

THEOREM 5.1. *There are universal constants $c, \delta > 0$ such that the following holds. Let $x_1, \ldots, x_m \in \{-1,1\}^n$ be a sequence of m independent uniformly random elements from $\{-1,1\}^n$. Then, if $m > (cn)^k \log(n^k)$, the probability that $X = \{x_1, \ldots, x_m\}$ contains a balanced k-wise independent distribution is at least $1 - \exp(-\delta^k m)$*

PROOF. By Theorem 3.1, $x_1, \ldots, x_m$ *does not* support a k-wise independent distribution if and only if there is a degree-k polynomial $f : \{-1,1\}^n \to \mathbb{R}$ with $\mathbb{E}[f] = 0$ such that $f(x_i) < 0$ for every $i \in [m]$.

For any fixed f, Theorem 4.1 gives that the probability that $f(x_i) < \tau^k$ for every x_i is at most $(1 - \tau^k)^m \leq \exp(-\tau^k m)$, where τ is the constant from Theorem 4.1. Thus, it is clear that any fixed f has a very small probability of

witnessing that $x_1, \ldots, x_m$ does not support a k-wise independent distribution.

To bound the probability that any f witnesses that x_1, ..., x_m supports a k-wise independent distribution, we construct a net of degree-k polynomials as follows: let $\mathcal{F}_\epsilon$ denote the set of degree-k polynomials $f : \{-1,1\} \to \mathbb{R}$ such that $\mathbb{E}[f] = 0$, $\mathrm{Var}[f] \leq 2$ and every coefficient of f is an integer multiple of ϵ.

We then have that $|\mathcal{F}_\epsilon| \leq (1/\epsilon)^{O(d_k)} = \exp(c_1 n^k \log 1/\epsilon)$ for some universal constant c_1. Then Theorem 4.1 and a union bound gives that the probability that there exists an $f \in \mathcal{F}_\epsilon$ such that $f(x_i) < \tau^k$ for every x_i, is bounded by

$$|\mathcal{F}_\epsilon|(1-\tau^k)^m \leq \exp(c_1 n^k \log(1/\epsilon) - \tau^k m) \leq \exp(-\tau^k m/2),$$

provided $m \geq 2c_1(n/\tau)^k \log(1/\epsilon)$.

Now, given an arbitrary degree-k polynomial f with $\mathbb{E}[f] = 0$, denote by $\tilde{f}$ the polynomial in $\mathcal{F}_\epsilon$ which is closest to f in ℓ_∞ norm. Then, if $||f - \tilde{f}||_\infty \leq \tau^k$ for every degree-k polynomial f, we would be done, since the existence of $f : \{-1,1\}^n \to \mathbb{R}$ such that $f(x_i) < 0$ for every x_i then implies the existence of $\tilde{f} \in \mathcal{F}_\epsilon$ such that $\tilde{f}(x_i) \leq f(x_i) + |\tilde{f}(x_i) - f(x_i)| < \tau^k$, which happens with probability at most $\exp(-\tau^k m/2) \leq \exp(-\delta^k m)$ for $\delta = \tau/2$.

We have the following easy bound on the distance $||f - \tilde{f}||_\infty$.

CLAIM 5.2. *For every f with $||f||_2 = 1$,*

$$||f - \tilde{f}||_\infty \leq \epsilon n^k,$$

provided this quantity is smaller than 1.

PROOF. Let f' be the result of rounding every coefficient of f to its nearest multiple of ϵ. Then, for any $x \in \{-1,1\}^n$,

$$|f(x) - f'(x)| = \left| \sum_{S \in D_k} (\hat{f}(S) - \hat{f'}(S)) \chi_S(x) \right| \leq \epsilon n^k,$$

where we used that $|D_k| \leq n^k$. It remains to show that $f' \in \mathcal{F}_\epsilon$, i.e., that $\mathrm{Var}[f'] \leq 2$. But this follows immediately since

$$\mathrm{Var}[f'] = ||f'||_2 \leq ||f||_2 + ||f - f'||_2 \leq 1 + ||f - f'||_\infty \leq 2$$

provided the bound on $||f - f'||_\infty \leq 1$. □

To finish the proof of Theorem 5.1, we thus conclude that in order to have $||f - \tilde{f}||_\infty \leq \tau^k$, it suffices to take

$$\epsilon = (\tau/n)^k.$$

Plugging this into the bound

$$m \geq 2c_1(n/\tau)^k \log(1/\epsilon)$$

we see that it suffices to take $m = (cn)^k \log(n^k)$ for c a constant depending only on τ, which in turn is a universal constant.

6. PAIRWISE INDEPENDENCE

In this section, we give our main theorem.

THEOREM 6.1. *There are constants $c, \delta > 0$ such that the following holds. Let $x_1, \ldots, x_m \in \{-1,1\}^n$ be a sequence of m independent uniformly random elements from $\{-1,1\}^n$. Then, if $m > cn^2$, the probability that $X = \{x_1, \ldots, x_m\}$ contains a balanced pairwise independent distribution is at least $1 - \exp(-\delta\sqrt{n})$.*

We get an immediate corollary.

COROLLARY 6.2. *There are constants $c, \delta > 0$ such that the following holds. Let $x_1, \ldots, x_s \in \{-1,1\}^t$ be a sequence of s independent uniformly random elements from $\{-1,1\}^t$. Let P be the predicate that accepts exactly the strings $(x_i)_{i=1}^s$. Then, assuming the UGC, if $s > ct^2$, the probability that P is approximation resistant is at least $1 - \exp(-\delta\sqrt{t})$.*

Before proceeding with the proof of Theorem 6.1, let us briefly describe the intuition behind it. The idea is to look at the convex hull K of the set of all ± 1 combinations of $x_1^{:\leq 2:}, \ldots, x_m^{:\leq 2:}$, and compare this to the sum $\overline{x} = x_1^{:\leq 2:} + \ldots + x_m^{:\leq 2:}$. By an application of Theorem 3.1, it suffices to prove that the latter sum lies strictly inside K with high probability. Intuitively, since $\overline{x}$ is a sum of m independent vectors with expected value $\mathbf{0}$ and length $\sqrt{d_2}$, the total length of $\overline{x}$ should be around $\sqrt{m \cdot d_2}$. On the other hand, K consists of all $[-1,1]$-valued linear combinations of $x_1^{:\leq 2:}, \ldots, x_m^{:\leq 2:}$ and as an easy consequence of hypercontractivity it will turn out that, in every direction v, each $x_i^{:\leq 2:}$ contributes a constant to the expected width of K in direction v. Thus one can hope that the size of K grows linearly in m so that if m is a sufficiently large multiple of d_2, K contains any vector of length $||\overline{x}|| \approx \sqrt{m \cdot d_2}$. It turns out that this is indeed the case, but in order to be able to show that the size of K grows linearly in *every* direction, we need to use the concentration inequality Theorem 2.8 for quadratic polynomials. It is this part which breaks down when one tries to repeat the same proof for k-wise independence in general—the necessary analogue of Theorem 2.8 is simply not true. We feel that this limitation to pairwise independence is a limitation of our proof rather than an inherent limitation in the problem, and that the analogue of Theorem 6.1 (where we require $m > (cn)^k$) should be true also for higher independence.

PROOF OF THEOREM 6.1. Let $m > cd_2$, where c is a constant that will be chosen sufficiently large. We will prove that, with probability at least $1 - \exp(-\delta\sqrt{n})$, for some $\delta > 0$, we have $\mathbf{0} \in \mathrm{Conv}(X^{:\leq 2:})$. By Theorem 3.1 this implies that X contains a pairwise independent distribution. This then implies Theorem 6.1, since $d_2 \leq n^2$.

Let

$$K = \left\{ \sum_{i=1}^m a_i x_i^{:\leq 2:} \ : \ |a_i| \leq 1 \right\},$$

and define

$$\overline{x} = \sum_{i=1}^m x_i^{:\leq 2:} \in \mathbb{R}^{d_2}.$$

Then, it suffices to prove that $\overline{x}$ lies in the interior of K, since if $\overline{x} = \sum_i a_i x^{:\leq 2:}$ with not all $a_i = 1$, we can rearrange and write $\mathbf{0}$ as the convex combination

$$\mathbf{0} = \sum_{i=1}^m \frac{1 - a_i}{\sum_j (1 - a_j)} x_i^{:\leq 2:} \in \mathrm{Conv}(X^{:\leq 2:}).$$

For a unit vector $v \in \mathbb{R}^{d_k}$, let

$$\mathrm{Width}(K, v) = \sup_{x \in K} \{\langle x, v \rangle\}$$

be the *width* of K in the direction v.

We will prove that, with high probability, the minimum width of K is larger than $||\overline{x}||$ (where $||\cdot||$ denotes the standard Euclidean norm in $\mathbb{R}^{d_k}$). In particular, we have the following two lemmas.

LEMMA 6.3. *There are constants $c_1 \in \mathbb{R}$, $c_2 > 0$ and $\delta_1 > 0$ such that, if $m > c_1 d_2$, the probability that*

$$\inf_v \text{Width}(K, v) < c_2 m \tag{2}$$

is at most $\exp(-\delta_1 m)$.

LEMMA 6.4. *There is a constant $\delta_2 > 0$ such that if $m \geq |D_2|$, the probability that*

$$||\overline{x}|| > 2\sqrt{md_2} \tag{3}$$

is at most $\exp(-\delta_2\sqrt{n})$.

Before proving the lemmas, let us see how they suffice to finish the proof of the theorem. Let $c = \max(c_1, (2/c_2)^2)$, and $m > cd_2$. Then by a union bound there is a δ such that with probability at least $1-\exp(-\delta\sqrt{n})$, neither Equation (2) nor Equation (3) holds, and we have

$$\inf_v \text{Width}(K, v) \geq c_2 m > 2\sqrt{md_2} \geq ||\overline{x}||.$$

This implies that $\overline{x}$ lies strictly inside K, as desired. Hence, if $m > cn^2 \geq c_0 d_2$, the probability that $\mathbf{0} \in \text{Conv}(X^{:\leq 2:})$ is at least $1 - \exp(-\delta\sqrt{n})$, and we are done. □

It remains to prove the two lemmas. We begin with Lemma 6.4 as this is the easier of the two.

PROOF OF LEMMA 6.4. Let

$$l = ||\overline{x}||^2 = \sum_{S \in D_2} \left(\sum_{i=1}^m \chi_S(x_i) \right)^2$$

be the squared length of $\overline{x}$. We can then view l as a degree 4 polynomial over $\{-1, 1\}^{mn}$. Our goal is to apply the concentration bound Theorem 2.7 to l. To be successful in this, we need that the variance $\text{Var}[l]$ is of a lower order than $\mathbb{E}[l]^2$. The expectation of l is easily seen to be $\mathbb{E}[l] = d_2 m$. To compute the variance of l, we compute

$$\begin{aligned} l^2 &= \sum_{S_1, S_2} \left(\sum_{i=1}^m \chi_{S_1}(x_i) \right)^2 \left(\sum_{i=1}^m \chi_{S_2}(x_i) \right)^2 \\ &= \sum_{S_1, S_2} \sum_{i_1, i_2, i_3, i_4 \in [m]} \chi_{S_1}(x_{i_1}) \chi_{S_1}(x_{i_2}) \chi_{S_2}(x_{i_3}) \chi_{S_2}(x_{i_4}). \end{aligned}$$

Define

$$A(S_1, S_2) = \sum_{i_1, i_2, i_3, i_4 \in [m]} \chi_{S_1}(x_{i_1}) \chi_{S_1}(x_{i_2}) \chi_{S_2}(x_{i_3}) \chi_{S_2}(x_{i_4}),$$

and let us analyze $\mathbb{E}[A(S_1, S_2)]$. If $S_1 \neq S_2$, the expected value of

$$\chi_{S_1}(x_{i_1}) \chi_{S_1}(x_{i_2}) \chi_{S_2}(x_{i_3}) \chi_{S_2}(x_{i_4})$$

is 0 unless $i_2 = i_1$ and $i_4 = i_3$. Hence for $S_1 \neq S_2$, we have

$$\mathbb{E}[A(S_1, S_2)] = \sum_{i_1, i_3} \mathbb{E}[\chi_{S_1}(x_{i_1})^2 \chi_{S_2}(x_{i_3})^2] = m^2,$$

since each term equals 1. Now let $S_1 = S_2 := S$, and consider the expected value of

$$\chi_S(x_{i_1}) \chi_S(x_{i_2}) \chi_S(x_{i_3}) \chi_S(x_{i_4}).$$

If for any $j \in [m]$ it is the case that only one of the i_k:s equal j, this expectation is 0, and otherwise the expectation is 1. Thus the only tuples (i_1, i_2, i_3, i_4) for which the expectation is not 0 are those where the values are paired up in the sense that $i_1 = i_2$ and $i_3 = i_4$, or $i_1 = i_3$ and $i_2 = i_4$, or $i_1 = i_4$ and $i_2 = i_3$. There are exactly $3m(m-1)+m$ ways to choose i_1, i_2, i_3, i_4 in such a paired way and hence in this case

$$\mathbb{E}[A(S, S)] = 3m(m-1) + m.$$

After these lengthy computations we thus find that

$$\mathbb{E}[l^2] = \sum_{S_1, S_2} \mathbb{E}[A(S_1, S_2)] = d_2^2 m^2 + 2d_2 m(m-1)$$

so that

$$\text{Var}[l] = 2d_2 m(m-1) \leq 2d_2 m^2,$$

Applying Theorem 2.7 to the polynomial $(l - \mathbb{E}[l])/\sqrt{\text{Var}[l]}$, we have

$$\begin{aligned} \Pr[||x|| > 2\sqrt{d_2 m}] &= \Pr[l - \mathbb{E}[l] > 3d_2 m] \\ &\leq \exp(-c(3d_2 m/\sqrt{\text{Var}[l]})^{1/2}) \leq \exp(-\delta_2 d_2^{1/4}), \end{aligned}$$

for $\delta_2 = c(9/2)^{1/4}$. Since $d_2 \geq n^2$, the lemma follows. □

We now move on to the proof of Lemma 6.3. By a standard argument the width is of K in any fixed direction is likely to be close to its expectation. Applying this to an ϵ-net of points we first prove that the maximum width of K is bounded and then proceed to establish also that the minimum is of the same order of magnitude.

LEMMA 6.5. *There are constants c_3 and $\tau > 0$ such that the following holds: for every $v \in \mathbb{R}^{d_2}$ with $||v|| = 1$, the probability that*

$$c_3 m \leq \text{Width}(K, v) \leq (1 + c_3)m$$

is at least $1 - \exp(-\tau m)$.

PROOF. Set $2c_3 = e^{-2}$. For $v \in \mathbb{R}^{d_2}$ with $||v|| = 1$, let $f_v : \{-1, 1\}^n \to \mathbb{R}$ be the corresponding degree-2 polynomial such that $f_v(x) = \langle v, x^{:\leq 2:} \rangle$.

By definition,

$$\text{Width}(K, v) = \max_{a \in [-1,1]^m} \sum_{i=1}^m a_i \left\langle v, x_i^{:\leq 2:} \right\rangle.$$

The maximum is clearly attained by setting

$$a_i = \text{sgn}\left(\left\langle v, x_i^{:\leq 2:} \right\rangle \right)$$

so that

$$\text{Width}(K, v) = \sum_{i=1}^m \left| \left\langle v, x_i^{:\leq 2:} \right\rangle \right| = \sum_{i=1}^m |f_v(x_i)|.$$

Applying Theorem 2.8 with $r = c_3\sqrt{m}$, the probability that $\sum_i |f_v(x_i)|$ deviates by more than $c_3 m$ from its expectation is at most $\exp(-\tau m)$ for some constant $\tau > 0$ (e.g., $\tau \leq \frac{c_3^2}{8e^2} - \frac{\ln 2}{m}$). But the expectation of $\sum_i |f_v(x_i)|$ equals $||f_v||_1 \cdot m$, which is trivially upper bounded by $||f_v||_2 \cdot m = m$, and by Theorem 2.6 lower bounded by $2c_3||f_v||_2 \cdot m = 2c_3 m$.

Hence, with probability at least $1 - \exp(-\tau m)$, we have

$$\begin{aligned} (||f_v||_1 - c_3)m &\leq \text{Width}(K, v) \leq (||f_v||_1 + c_3)m \\ c_3 m &\leq \text{Width}(K, v) \leq (1 + c_3)m. \end{aligned}$$

□

We now prove the lower bound on the minimum width of K.

PROOF OF LEMMA 6.3. Let $V = \{v_1, \ldots, v_L\}$ be an ϵ-net of the unit sphere in $\mathbb{R}^{d_2}$, i.e., a set of vectors such that, for every $v \in \mathbb{R}^{d_2}$ with $||v|| = 1$, there is a vector $v_i \in V$ such that $\langle v, v_i \rangle \geq 1 - \epsilon$. As stated in Theorem 2.1 such a set can be constructed of size at most $L = (5/\epsilon)^{|d_2|}$.

For any $v_i \in V$, Lemma 6.5 tells us that

$$c_3 m \leq \text{Width}(K, v_i) \leq (1 + c_3)m$$

except with probability at most $\exp(-\tau m)$. By a union bound, these inequalities then hold for every $v_i \in V$ except with probability

$$L \exp(-\tau m) = \exp(-\tau m + \ln(5/\epsilon) d_2) = \exp(-\tau m/2),$$

provided m is a sufficiently large multiple of $d_2 \cdot \ln(1/\epsilon)$.

Let $W_{\max} = \sup_{||v||=1} \text{Width}(K, v)$. We now prove that $W_{\max}$ is small.

For any $w \in \mathbb{R}^{d_2}$ with $||w|| = 1$, we can write $w = (1 - \epsilon')v_i + \sqrt{1 - (1-\epsilon')^2} w'$ for some $\epsilon' \leq \epsilon$, $v_i \in V$ and unit vector w'. We then have for any $u \in K$

$$\begin{aligned} \langle u, w \rangle &= (1 - \epsilon') \langle u, v_i \rangle + \sqrt{\epsilon'(2 - \epsilon')} \langle u, w' \rangle \\ &\leq \text{Width}(K, v_i) + \sqrt{2\epsilon}\, \text{Width}(K, w') \\ &\leq (1 + c_3)m + \sqrt{2\epsilon} W_{\max}. \end{aligned}$$

Taking the supremum over all $u \in K$ and unit vectors $w \in \mathbb{R}^{d_2}$, we obtain

$$\begin{aligned} W_{\max} &\leq (1 + c_3)m + \sqrt{2\epsilon} W_{\max} \\ W_{\max} &\leq \frac{1 + c_3}{1 - \sqrt{2\epsilon}} m \leq (1 + 2c_3)m, \end{aligned}$$

provided ϵ is chosen sufficiently small compared to c_3.

Having established that K is not too wide in any direction we can now prove that it is not too narrow completing the proof of Lemma 6.3.

We have, again for any $w = (1-\epsilon')v_i + \sqrt{\epsilon'(2-\epsilon')} w'$ and $u \in K$,

$$\begin{aligned} \langle u, w \rangle &= (1 - \epsilon') \langle u, v_i \rangle + \sqrt{\epsilon'(2 - \epsilon')} \langle u, w' \rangle \\ &\geq (1 - \epsilon) c_3 m - \sqrt{2\epsilon}\, \text{Width}(K, w') \\ &\geq ((1 - \epsilon)c_3 - \sqrt{2\epsilon}(1 + 2c_3))m \geq c_3 m/2, \end{aligned}$$

again provided ϵ is sufficiently small compared to c_3.

Hence, with probability at least $1 - \exp(-\delta m)$, we have $\inf_{||v||=1} \text{Width}(K, v) \geq c_3 m/2 := c_2 m$, provided that m is a sufficiently large multiple $c_1 d_2$ of d_2. □

7. A LOWER BOUND FOR RANDOM SUPPORT SIZE

In this section we give a lower bound on the threshold for randomly supported independence.

THEOREM 7.1. *There exists a universal constant δ such that the following holds. Let $x_1, \ldots, x_m$ be a sequence of m independent uniformly random samples from $\{-1, 1\}^n$. Then, if $m < \frac{n^k}{6^{k^2}}$, the probability that $x_1, \ldots, x_m$ can support a balanced k-wise independent distribution is at most*

$$\exp(-\Omega(\delta^k n)).$$

PROOF. We will prove that, if $m \leq \frac{n^k}{6^{k^2}}$, then with high probability $x_1^{:\leq k:}, \ldots, x_m^{:\leq k:}$ are linearly independent. In particular, this implies that any convex combination of $x_1^{:\leq k:}$, …, $x_m^{:\leq k:}$ is non-zero, so that, by Theorem 3.1, $x_1, \ldots, x_m$ does not support a k-wise independent distribution.

The main component of the proof is the following lemma.

LEMMA 7.2. *Let $m \leq \frac{n^k}{6^{k^2}}$, and let $y_1, \ldots, y_m \in \mathbb{R}^{d_k}$ be m arbitrary points. Then, the probability that a uniformly random point $x \in \{-1, 1\}^n$ has $x^{:\leq k:}$ lying in the space spanned by $y_1, \ldots, y_m$ is at most $\exp\left(-\frac{n}{2k2^k}\right)$.*

Before proving the lemma we finish the proof of the theorem. Let $m = \frac{n^k}{6^{k^2}}$, and let $x_1, \ldots, x_m$ be m uniformly random points of $\{-1, 1\}^n$. Using Lemma 7.2, we conclude that the probability that $x_1^{:\leq k:}, \ldots, x_m^{:\leq k:}$ are linearly independent is at least

$$1 - m \exp\left(-\frac{n}{2k2^k}\right) \geq 1 - \exp(-\Omega(\delta^k n)),$$

for $\delta = 1/3$ (say), which proves Theorem 7.1. □

Next, we turn to the proof of the lemma.

PROOF OF LEMMA 7.2. Let $S \subseteq \mathbb{R}^{d_k}$ be the space spanned by the vectors $y_1, \ldots, y_m$. Then S has dimension at most m and hence is determined by at least $d_k - m$ linearly independent equations $v_1, \ldots, v_{d_k - m} \in \mathbb{R}^{d_k}$ such that $y \in S$ iff $\langle v_i, y \rangle = 0$ for every $i \in [d_k - m]$. Equivalently, for $x \in \{-1, 1\}^n$, we have $x^{:\leq k:} \in S$ iff $v_i(x) = 0$ for every i, where we again interpret v_i as a degree-k polynomial. We will prove that only an exponentially small fraction of all points $x \in \{-1, 1\}^n$ satisfy these conditions.

In what follows, we explicitly refer to d_k as a function of n, i.e.,

$$d_k(n) := \sum_{i=1}^{k} \binom{n}{i} \geq \left(\frac{n}{k}\right)^k,$$

Let $T(n, m)$ be the maximum possible number of solutions $x \in \{-1, 1\}^n$ to a system of at least $d_k(n) - m$ linearly independent degree-k polynomial equations $v_1(x) = 0$, …, $v_{d_k(n)-m}(x) = 0$. We will prove that

$$T(n, m) \leq (2^k - 1)^{n/k} \cdot \exp(k m^{1/k}). \tag{4}$$

If $d_k(n) \leq m$ so that $n \leq m^{1/k} k$, we have the trivial bound $T(n, m) \leq 2^n \leq \exp(k m^{1/k})$, so let $d_k(n) > m$ and assume inductively that Equation (4) holds for all $n' < n$. Assume that there is a v_i which has degree exactly k (if all v_i have degree at most $k - 1$, we would get an even better bound). Without loss of generality, we can take v_1 to have degree exactly k, and having $[k]$ as a non-zero coefficient, i.e., $\hat{v_1}([k]) \neq 0$.

Next, eliminate (by standard Gaussian elimination) all coordinates S with $S \cap [k] \neq \emptyset$. As there are exactly $d_k(n) - d_k(n-k)$ such values of S, the resulting system has at least $(d_k(n) - m) - (d_k(n) - d_k(n-k)) = d_k(n-k) - m$ equations, and hence has at most $T(n-k, m)$ solutions. Let us, for each such solution $x^* \in \{-1, 1\}^{n-k}$, consider the number of ways of extending it to a solution for the original system. Plugging in x^* in the equation $v_1(x) = 0$, this equation becomes an equation of the form

$$p(x_{[k]}) = 0,$$

for some function $p : \{-1,1\}^k \to \mathbb{R}$. Furthermore, the function p is not identically zero, since $\hat{p}([k]) \neq 0$. This implies that the number of ways of extending x^* is at most $2^k - 1$, and hence we have

$$T(n,m) \leq (2^k-1) \cdot T(n-k,m) \leq (2^k-1)^{n/k} \cdot \exp(km^{1/k}).$$

Thus, the probability that $x^{:\leq k:}$ lies inside S for a uniformly random point $x \in \{-1,1\}^n$ is at most

$$\begin{aligned}(2^k-1)^{n/k}\exp(km^{1/k})/2^n &= (1-2^{-k})^{n/k}\exp(km^{1/k}) \\ &\leq \exp\left(-\frac{n}{k2^k} + km^{1/k}\right).\end{aligned}$$

Plugging in $m \leq \frac{n^k}{6^{k^2}}$, we have $km^{1/k} \leq \frac{n}{2k2^k}$, and the lemma follows. □

8. APPROXIMATING A RANDOM PREDICATE

In this section we let P be a predicate constructed by randomly choosing $O(t^2/\log t)$ t-bit strings and making these be the inputs accepted by P. We have the following theorem.

THEOREM 8.1. *There is a constant $c > 0$ such that the following is true. Suppose $s \leq ct^2/\log t$ and suppose $P : \{-1,1\}^t \mapsto \{0,1\}$ is a predicate chosen randomly among all predicates that accept s inputs. Then, with probability $1 - \frac{1}{t}$, P is not approximation resistant.*

In the analysis we assume that the s strings accepted by P are chosen with replacement and hence are independent. Since the strings are distinct with probability $1 - O(t^4 2^{-t})$ this is sufficient to prove the theorem.

As discussed in Section 2, P can be represented by a multilinear polynomial and in this section the quadratic part, denoted by $P^{=2}$, is of special importance.

The following lemma is a special case of Theorem 4.9 (using $C = 0$) of [11].

LEMMA 8.2. *Suppose $P^{=2}(y) > 0$ for any $y \in P^{-1}(1)$, then P is not approximation resistant.*

The key technical lemma to apply the above lemma is the following.

LEMMA 8.3. *Suppose P is constructed as in the hypothesis of Theorem 8.1, then for any $y \in P^{-1}(1)$,*

$$\Pr[P^{=2}(y) \leq 0] \leq t^{-3}.$$

Using an application of the union bound it is easy to see that Lemma 8.2 and Lemma 8.3 jointly imply Theorem 8.1 and thus all we need to do is to establish Lemma 8.3.

PROOF OF LEMMA 8.3. $P^{=2}$ is the quadratic form

$$P^{=2}(x) = \sum_{i<j} \hat{P}_{ij} x_i x_j$$

where

$$\hat{P}_{ij} = 2^{-t} \sum_{z \in P^{-1}(1)} z_i z_j.$$

Now for $y \in P^{-1}(1)$ we see that

$$\begin{aligned}P^{=2}(y) &= 2^{-t} \sum_{i<j} \sum_{z \in P^{-1}(1)} z_i z_j y_i y_j. \\ &= 2^{-t}\left(\binom{t}{2} + \sum_{\substack{z \in P^{-1}(1) \\ z \neq y}} \sum_{i<j} z_i z_j y_i y_j\right) \quad (5)\end{aligned}$$

The sum in Equation (5) is of the form $\sum_z P_y(z)$ where P_y is a quadratic polynomial such that $\mathbb{E}[P_y(z)] = 0$ and $\mathbb{E}[(P_y(z))^2] = \binom{t}{2}$. As we are summing P_y at $s-1$ random points we have, if $r \leq 2e\sqrt{s-1}$, by Theorem 2.8, that

$$\Pr\left[\left|\sum_z P_y(z)\right| \geq r\sqrt{(s-1)\binom{t}{2}}\right] \leq \exp(-\Omega(r^2)).$$

Setting $r = \sqrt{\binom{t}{2}/(s-1)}$, this implies, for $s = \omega(t)$, that

$$\Pr\left[\left|\sum_z P_y(z)\right| \geq \binom{t}{2}\right] \leq \exp(-\Omega(t^2/s)) \leq 1/t^3 \quad (6)$$

for an appropriately chosen $s = \Theta(t^2/\log t)$ and the proof of the lemma is complete. □

9. CONCLUDING REMARKS

Assuming the UGC we have established rather tight bounds on the density at which a random predicate is likely to become approximation resistant. This indicates that approximation resistance is the typical property of a predicate and only very sparse or very special predicates can be efficiently approximated in a non trivial way.

It is difficult not to view this paper as yet another reason that we must, if possible, settle the Unique Games Conjecture in the close future. Another road ahead is of course to prove the results without the UGC but it is not obvious that this is significantly easier.

On a detailed technical level, although our results are rather tight we have two annoying logarithmic gaps that should be closed.

We feel that it is likely that $O(n^k)$ random points are sufficient to support a k-wise independent distribution with good probability. For the case of the density at which a random predicate becomes approximation resistance we feel less convinced of the correct answer but our inclination is to believe that the correct answer is $\Theta(t^2)$.

9.1 Acknowledgments

We are grateful to Elchanan Mossel and Ryan O'Donnell for interesting discussions, and to Ryan also for the proof of Theorem 4.3 as discussed in Section 4. We also acknowledge the financial support by Swedish Research Council Project Number 50394001 and ERC Advanced investigator grant 226203.

10. REFERENCES

[1] S. Arora, S. Khot, A. Kolla, D. Steurer, M. Tulsiani, and N. K. Vishnoi. Unique games on expanding constraint graphs are easy. In *ACM Symposium on Theory of Computing (STOC)*, pages 21–28, 2008.

[2] P. Austrin. *Conditional Inapproximability and Limited Independence*. PhD thesis, KTH – Royal Institute of Technology, 2008.

[3] P. Austrin and E. Mossel. Approximation Resistant Predicates From Pairwise Independence. In *IEEE Conference on Computational Complexity (CCC)*, 2008.

[4] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102(1):159–182, 1975.

[5] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier*, 20:335–402, 1970.

[6] M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for unique games. In *ACM Symposium on Theory of Computing (STOC)*, pages 205–214, 2006.

[7] I. Dinur, E. Friedgut, G. Kindler, and R. O'Donnell. On the Fourier tails of bounded functions over the discrete cube. In *ACM Symposium on Theory of Computing (STOC)*, pages 437–446, 2006.

[8] Z. Füredi. Random Polytopes in the d-Dimensional Cube. *Discrete Comput. Geom.*, 1:315–319, 1986.

[9] M. X. Goemans and D. P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the ACM*, 42:1115–1145, 1995.

[10] V. Guruswami, D. Lewin, M. Sudan, and L. Trevisan. A tight characterization of NP with 3 query PCPs. In *Proceedings of 39th Annual IEEE Symposium on Foundations of Computer Science*, pages 8–17, Palo Alto, 1998. IEEE.

[11] G. Hast. *Beating a Random Assignment – Approximating Constraint Satisfaction Problems*. PhD thesis, KTH – Royal Institute of Technology, 2005.

[12] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[13] J. Håstad. Every 2-CSP Allows Nontrivial Approximation. In *ACM Symposium on Theory of Computation (STOC)*, pages 740–746, 2005.

[14] J. Håstad. On the approximation resistance of a random predicate. In *APPROX-RANDOM*, pages 149–163, 2007.

[15] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, 1997.

[16] S. Khot. On the power of unique 2-prover 1-round games. In *ACM Symposium on Theory of Computing (STOC)*, pages 767–775, 2002.

[17] M. Kochol. Constructive approximation of a ball by polytopes. *Math. Slovaca*, 44(1):99–105, 1994.

[18] E. Mossel. Gaussian bounds for noise correlation of functions. arXiv Report math/0703683v3, 2007.

[19] A. Samorodnitsky and L. Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *ACM Symposium on Theory of Computing (STOC)*, pages 191–199, 2000.

[20] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influence of variables, and PCPs. In *ACM Symposium on Theory of Computing (STOC)*, pages 11–20, 2006.

[21] L. Trevisan. Approximation algorithms for unique games. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 197–205, 2005.

[22] U. Zwick. Approximation Algorithms for Constraint Satisfaction Problems Involving at Most Three Variables Per Constraint. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 1998.

APPENDIX

A. PROOFS OF CONCENTRATION BOUNDS

PROOF OF THEOREM 2.7. Set $p = t^{2/d} \cdot \frac{1}{e}$. By Markov's inequality, we have

$$\Pr[|f| > t] = \Pr[|f|^p > t^p] \leq \frac{||f||_p^p}{t^p}. \tag{7}$$

Now, since $t > e^{d/2}$, p is at least 1. This implies, using Theorem 2.3 for $p > 2$ and the monotonicity of ℓ_p norms for $1 \leq p \leq 2$, that

$$||f||_p \leq \sqrt{p}^d ||f||_2 = te^{-d/2}.$$

Plugging this into Equation (7) we get

$$\Pr[|f| > t] \leq \left(\frac{te^{-d/2}}{t}\right)^p = \exp(-pd/2) = \exp\left(-\frac{d}{2e}t^{2/d}\right).$$

□

PROOF OF THEOREM 2.8. By Markov's inequality and the standard Chernoff method, we have

$$\Pr\left[\sum_{i=1}^m f(x_i) - m\,\mathbb{E}[f] > r\sqrt{m}\right] \leq \frac{\prod_{i=1}^m \mathbb{E}[\exp(\lambda f(x_i))]}{\exp(\lambda m\,\mathbb{E}[f] + \lambda r\sqrt{m})}. \tag{8}$$

We use the Taylor expansion of $\exp(x) = \sum_{k=0}^\infty x^k/k!$ and Theorem 2.3 to bound the expression $\mathbb{E}[\exp(\lambda f(x_i))]$:

$$\begin{aligned}
\mathbb{E}[\exp(\lambda f(x_i))] &= \sum_{k=0}^\infty \frac{\mathbb{E}\left[(\lambda f(x_i))^k\right]}{k!} \\
&\leq 1 + \lambda\,\mathbb{E}[f] + \sum_{k=2}^\infty \frac{(\lambda k)^k}{k!} \leq 1 + \lambda\,\mathbb{E}[f] + \sum_{k=2}^\infty \frac{(\lambda k)^k}{(k/e)^k} \\
&= 1 + \lambda\,\mathbb{E}[f] + \frac{(\lambda e)^2}{1-\lambda e} \leq \exp(\lambda\,\mathbb{E}[f] + 2\lambda^2 e^2),
\end{aligned}$$

where the last two steps assume that λ is small enough so that $\lambda \leq (2e)^{-1}$. Hence, the bound in Equation (8) becomes

$$\frac{\prod_{i=1}^m \mathbb{E}[\exp(\lambda f(x_i))]}{\exp(\lambda m\,\mathbb{E}[f] + \lambda r\sqrt{m})} \leq \exp(2\lambda^2 e^2 m - \lambda r\sqrt{m})$$

This is minimized for $\lambda = \frac{r\sqrt{m}}{4e^2 m} = \frac{r}{4e^2\sqrt{m}}$ (the bound $r < 2e\sqrt{m}$ guarantees that the assumption $\lambda \leq (2e)^{-1}$ is satisfied). Plugging in this value of λ gives the bound

$$\begin{aligned}
\Pr\left[\sum_{i=1}^m f(x_i) - m\,\mathbb{E}[f] > r\sqrt{m}\right] &\leq \exp\left(-\frac{r^2 m}{8e^2 m}\right) \\
&= \exp\left(-\frac{r^2}{8e^2}\right).
\end{aligned}$$

The bound on $\Pr\left[\sum_{i=1}^m f(x_i) - m\,\mathbb{E}[f] < -r\sqrt{m}\right]$ follows by applying the first inequality to the function $-f$. That the bounds hold also when f is replaced by $|f|$ follows by the fact that the only property of f that was used was that its moments are bounded, and taking absolute value does not change moments. □