RANDOMLY SUPPORTED INDEPENDENCE AND RESISTANCE*

PER AUSTRIN† AND JOHAN HÅSTAD‡

Abstract. We prove that for any positive integers q and k, there is a constant $c_{q,k}$ such that a uniformly random set of $c_{q,k}n^k \log n$ vectors in $[q]^n$ 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 $c_{q,k}n^k$. Using a recent result by Austrin and Mossel this shows that a predicate on t bits, chosen at random among predicates accepting $c_{q,2}t^2$ 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, $c'_{q,k}$, such that a randomly selected set of cardinality $c'_{q,k}n^k$ 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 Boolean inputs accepting at least $(32/33) \cdot 2^t$ inputs is approximation resistant.

The results extend from balanced distributions to arbitrary product distributions.

Key words. k-wise Independence, Constraint Satisfaction, Approximation Resistance

AMS subject classifications. 68Q17 (primary), 68W25 (secondary)

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 (one can allow larger domain sizes but for simplicity in this motivational discussion we stay with the Boolean domain). 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. The reason we say "almost" is that 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 [13]. The famous approximation algorithm of Goemans and Williamson [10] 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 [14].

*Research supported by Swedish Research Council Project Number 50394001 and ERC Advanced investigator grant 226203. A preliminary version of this paper appeared in the proceedings of STOC 2009.

†University of Toronto, Toronto, Canada (austrin@cs.toronto.edu).

‡KTH – Royal Institute of Technology, Stockholm, Sweden (johanh@kth.se).

Zwick [25] 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 [11, 21] and in particular Hast [12] 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 *hereditarily 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. [11] but implies $NAE(x_2, x_3, x_4)$ (the "not-all-equal" predicate) which admits a nontrivial approximation algorithm, see for instance [25].

As a generic positive result Hast [12] 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 [17]. 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 [23, 7, 2, 19, 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.

Assuming the UGC, Samorodnitsky and Trevisan [22] 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 [15] 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 on the form $2^r - 1$, [15] improves the value of s to $2^t/t$ but this is shown to be the lower limit of what can be obtained using the predicates of Samorodnitsky and Trevisan.

Austrin and Mossel [4], using the machinery of Mossel [20] 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 result they established, without assumptions on the form of t, that there are predicates that accept $t + o(t)$ inputs which satisfy this property. Furthermore if the famous *Hadamard 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 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 over $[q]^n$ for some alphabet size q, establishing the following result.

THEOREM 1.1 (informal). *There are absolute constants $c_{q,k}$ such that if we pick $c_{q,k}n^k \log n$ random points in $[q]^n$, 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 are absolute constants $c_{q,2}$ such that if we pick $c_{q,2}n^2$ random points in $[q]^n$, 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 over $\{0,1\}^n$, i.e., the case $k = 1$ and $q = 2$, a sharp bound of $2n$ on the threshold is already known by an elegant result by Füredi [9].

The bounds for the case $k \leq 2$ are asymptotically tight: in Theorem 7.1 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 are absolute constants $c_{q,2}$ such that if we pick a random predicate $P : [q]^t \to \{0,1\}$ on t inputs which accepts $c_{q,2}t^2$ of the q^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 not automatically give an answer to the question when a predicate is approximation resistant. Here we get an almost tight result by showing in Theorem 8.1 that, for some constant $c_q > 0$, a predicate that accepts a random set of size $c_q t^2/\log t$ is likely to admit a nontrivial approximation algorithm. 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 (informal). *There are constants $c_q > 0$ such that any subset of size $(1 - c_q^k)q^n$ of $[q]^n$ supports a k-wise independent distribution.*

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

COROLLARY 1.5. *Any predicate on t Boolean inputs that accepts at least $(32/33) \cdot 2^t$ 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 $P \neq NP$ [12], and that any predicate accepting more than $2^t(1 - (2t)^{-1})$ inputs is resistant assuming the UGC [15].

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 [12] 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 $q = 2$ and $k = 2$, the correct value is strictly larger than $13/16$ as we establish in Section 4.

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.

All results extend to arbitrary product distributions over $[q]^n$, not just the uniform one. We refer to the respective theorem statements for precise details.

2. Preliminaries. Let Ω be a finite set and let μ be a probability distribution on Ω which has full support in that $\mu(x) > 0$ for any $x \in \Omega$. The following notation is used throughout the paper.

- $(\Omega^n, \mu^{\otimes n})$ denotes the product space $\Omega \times \ldots \times \Omega$, endowed with the product distribution.
- $\alpha(\mu) = \min\{\, \mu(x) \,:\, x \in \Omega, \mu(x) > 0 \,\} = \min\{\mu(x) : x \in \mathrm{Supp}(\mu)\}$ denotes the minimum non-zero probability of any atom in Ω under the distribution μ.
- $L^2(\Omega, \mu)$ denotes the space of functions from Ω to $\mathbb{R}$. We define the inner product on $L^2(\Omega, \mu)$ by $\langle f, g \rangle_\mu = \mathbb{E}_{x \in (\Omega, \mu)}[f(x)g(x)]$, and ℓ_p norm by $||f||_p = (\mathbb{E}_{x \in (\Omega, \mu)}[|f(x)|^p])^{1/p}$. The ℓ_∞ norm of f is defined by $||f||_\infty = \max_{\mu(x)>0} |f(x)|$.

We generally use $\mathcal{U}$ to denote the uniform distribution. So e.g. $(\{-1,1\}^n, \mathcal{U})$ denotes the Boolean hypercube endowed with the uniform distribution. 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 $f, g \in L^2(\Omega, \mu)$. Then

$$\langle f, g \rangle_\mu \leq ||f||_p \cdot ||g||_q$$

For a probability distribution η over Ω^n (not necessarily a product distribution) and subset $S \subseteq [n]$ of coordinates, we denote by η_S the *marginal distribution* of η on the coordinates in S (i.e., the distribution on $\Omega^{|S|}$ induced by η by only looking at the coordinates in S). A distribution η over Ω^n is *k-wise independent* if, for every $S \subseteq [n]$ with $|S| = k$, it holds that η_S is a product distribution. If, additionally, each such η_S is the uniform distribution over Ω^k, we say that η is *balanced* k-wise independent.

For vectors $u, v \in \mathbb{R}^n$, we denote by $\langle u, v \rangle = \sum_{i=1}^n u_i v_i$ the standard inner product in $\mathbb{R}^n$. 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 hopefully is 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 also need the following standard result on small ϵ-nets of the unit sphere (see e.g. [18]):

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. Fourier Decomposition. In this subsection we recall some background in Fourier analysis that is used in the paper.

Let q be a positive integer (not necessarily a prime power), and let (Ω, μ) be a finite probability space with $|\Omega| = q$ and full support, i.e., $\mathrm{Supp}(\mu) = \Omega$. Let $\chi_0, \ldots, \chi_{q-1} : \Omega \to$

$\mathbb{R}$ be an orthonormal basis for the space $L^2(\Omega, \mu)$ w.r.t. the scalar product $\langle \cdot, \cdot \rangle_\mu$. Furthermore, we require that this basis has the property that $\chi_0 = \mathbf{1}$, i.e., the function that is identically 1 on every element of Ω. As we work with product spaces the following definition is useful for us.

DEFINITION 2.2. *A* multi-index *is a vector $\sigma \in \mathbb{Z}_q^n$, for some q and n. The* support *of a multi-index σ is $S(\sigma) = \{ i : \sigma_i > 0 \} \subseteq [n]$.*

For readability we slightly abuse notation and treat a multi-index as a set, e.g. writing $|\sigma|$ instead of $|S(\sigma)|$, $i \in \sigma$ instead of $i \in S(\sigma)$, and so on.

For a multi-index σ, define $\chi_\sigma : \Omega^n \to \mathbb{R}$ as $\bigotimes_{i \in [n]} \chi_{\sigma_i}$, i.e.,

$$\chi_\sigma(x_1, \ldots, x_n) = \prod_{i \in [n]} \chi_{\sigma_i}(x_i).$$

It is well-known and easy to check that the functions $\{\chi_\sigma\}_{\sigma \in \mathbb{Z}_q^n}$ form an orthonormal basis for the product space $L^2(\Omega^n, \mu^{\otimes n})$. Thus, every function $f \in L^2(\Omega^n, \mu^{\otimes n})$ can be written as

$$f(x) = \sum_{\sigma \in \mathbb{Z}_q^n} \hat{f}(\sigma) \chi_\sigma(x),$$

where $\hat{f} : \mathbb{Z}_q^n \to \mathbb{R}$ is defined by $\hat{f}(\sigma) = \langle f, \chi_\sigma \rangle_{\mu^{\otimes n}}$. The most basic properties of $\hat{f}$ are summarized by Fact 2.3, which is an immediate consequence of the orthonormality of $\{\chi_\sigma\}_{\sigma \in \mathbb{Z}_q^n}$.

FACT 2.3. *We have*

$$\mathbb{E}[fg] = \sum_\sigma \hat{f}(\sigma)\hat{g}(\sigma) \qquad \mathbb{E}[f] = \hat{f}(\mathbf{0}) \qquad \mathrm{Var}[f] = \sum_{\sigma \neq \mathbf{0}} \hat{f}(\sigma)^2.$$

We refer to the transform $f \mapsto \hat{f}$ as the Fourier transform, and $\hat{f}$ as the Fourier coefficients of f. We remark that the article "the" is somewhat inappropriate, since the transform and coefficients in general depend on the choice of basis $\{\chi_i\}_{i \in \mathbb{Z}_q}$. However, we always work with some fixed (albeit arbitrary) basis, and hence there should be no ambiguity in referring to the Fourier transform as if it were unique. Furthermore, many of the important properties of $\hat{f}$ are actually basis-independent.

We say that a polynomial $f \in L^2(\Omega^n, \mu^{\otimes n})$ has degree d if $\hat{f}(\sigma) = 0$ for every σ with $|\sigma| > d$. We let $f^{=d}$ denote the part of f that is of degree exactly d. Note that in this notation an arbitrary function is a polynomial of degree n.

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

$$D_k := D_k(n) = \{ \sigma \in \mathbb{Z}_q^n \mid 1 \leq |\sigma| \leq k \},$$

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

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 2.4. *Given a string $x \in \Omega^n$, we define $x^{:\leq k:}$ as*

$$x^{:\leq k:} = (\chi_\sigma(x))_{\sigma \in D_k} \in \mathbb{R}^{d_k},$$

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 $X \subseteq \Omega^n$, we use $X^{:\leq k:} \subseteq \mathbb{R}^{d_k}$ to denote the set $\{ x^{:\leq k:} \mid x \in X \}$.

Note that every $v \in \mathbb{R}^{d_k}$ is in 1–1 correspondence with a degree-k polynomial $f_v \in L^2(\Omega^n, \mu^{\otimes n})$ with $\mathbb{E}[f_v] = 0$, defined by $f_v(x) = \left\langle v, x^{:\leq k:} \right\rangle$ for every $x \in \Omega^n$ (i.e., we interpret v as giving the Fourier coefficients of f_v).

Another fact which is sometimes useful is the following trivial bound on the ℓ_∞ norm of χ_σ (recall that $\alpha(\mu)$ is the minimum non-zero probability of any atom in μ which we assume to be fully supported).

FACT 2.5. *Let $(\Omega^n, \mu^{\otimes n})$ be a product space with Fourier basis $\{\chi_\sigma\}_{\sigma \in \mathbb{Z}_q^n}$. Then for any $\sigma \in \mathbb{Z}_q^n$,*

$$||\chi_\sigma||_\infty \leq \alpha(\mu)^{-|\sigma|/2}.$$

Proof. It is clearly enough to prove this for any basis function χ_i and this case follows from that by the orthonormality of these functions since

$$1 = \sum_x \chi_i^2(x)\mu(x) \geq \chi_i^2(x_0)\mu(x_0)$$

for any fixed x_0. □

2.2. Norm Inequalities. The main analytic tool in all our upper bounds are "Khinchin type" inequalities for low degree polynomials, i.e., the fact that the ℓ_p norms of such polynomials are related to within a constant factor. Such bounds follow in turn from *hypercontractivity* estimates for such functions. For instance, a well-known consequence of the famous *Hypercontractivity Theorem* [6, 5] can be stated as follows.

THEOREM 2.6. *Let $f \in L^2(\{-1,1\}^n, \mathcal{U})$ be a degree-d polynomial. Then, for every $1 \leq q < p \leq \infty$, it holds that*

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

Recall that the classic Khinchin Inequality states that $||f||_p \leq C_p||f||_2$ for degree-1 polynomials and some constant C_p depending only on p. Using the recent sharp hypercontractivity estimates of Wolff [24], we have the following generalization to arbitrary finite probability spaces.

THEOREM 2.7 ([24]). *Let (Ω, μ) be a finite probability space in which the minimum non-zero probability is $\alpha = \alpha(\mu) \leq 1/2$. Then for $p \geq 2$, every degree-d polynomial $f \in L^2(\Omega^n, \mu^{\otimes n})$ satisfies*

$$||f||_p \leq C_p(\alpha)^{d/2}||f||_2.$$

Here C_p is defined by

$$C_p(\alpha) = \frac{A^{1/p'} - A^{-1/p'}}{A^{1/p} - A^{-1/p}}$$

where $A = (1-\alpha)/\alpha$ and $1/p + 1/p' = 1$. The value at $\alpha = 1/2$ is taken to be the limit of the above expression as $\alpha \to 1/2$, i.e., $C_p(1/2) = p - 1$.

As the quantity $C_p(\alpha)$ is somewhat ungainly to work with, the following upper bounds can be convenient.

FACT 2.8. *For every $\alpha \in (0, 1/2]$ and $p \geq 2$ we have $C_p(\alpha) \leq \frac{p}{2\alpha}$ and $C_3(\alpha) \leq \left(\frac{4}{\alpha}\right)^{1/3}$.*

Proof. The first bound is proven in [3] as Corollary 7.1.4. For the second bound, note that $p' = 3/2$ and hence

$$C_3(\alpha) = \frac{A^{2/3} - A^{-2/3}}{A^{1/3} - A^{-1/3}} = A^{1/3} + A^{-1/3}.$$

Now for any $x > 0$ we have

$$(x + x^{-1})^3 \leq 4(x^3 + x^{-3}).$$

This follows as the difference between the two sides is

$$3x^3 + 3x^{-3} - 3x - 3x^{-1} = 3(x^2 - 1)(x - x^{-3})$$

which clearly is nonnegative. Applying this with $x = A^{1/3}$ we get

$$A^{1/3} + A^{-1/3} \leq (4(A + A^{-1}))^{1/3} \leq (4(A + 1))^{1/3} = \left(\frac{4}{\alpha}\right)^{1/3}$$

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 2.7. It does however follow from a classic "duality" argument.

THEOREM 2.9. *Let f be a random variable. If f satisfies $||f||_p \leq C||f||_2$ for some constants $p > 2$ and C, then*

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

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,

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

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

Combined with Theorem 2.7 and Fact 2.8 and taking $p = 3$, this implies the following bound.

COROLLARY 2.10. *Let $f \in L^2(\Omega^n, \mu^{\otimes n})$ be a degree-d polynomial. Then*

$$||f||_2 \leq \left(\frac{4}{\alpha(\mu)}\right)^{d/2} ||f||_1.$$

In some cases, stronger relations between the norms of f are possible than can be obtained by going via hypercontractivity. The following estimate for the uniform Boolean hypercube in the case $p = 2$, $q = 4$, and $d = 2$ (i.e., quadratic polynomials) is sometimes useful (note that Theorem 2.6 in this case gives the constant $1/3 = 81^{-1/4}$).

THEOREM 2.11. *Let $f \in L^2(\{-1, 1\}^n, \mathcal{U})$ 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 [16], 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 multiplicity 1, the second type with multiplicity 6 and the last with multiplicity 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. □

2.3. Concentration Bounds. It is known that hypercontractivity implies good concentration bounds for low-degree polynomials (see e.g. [8]). We need the following two results, and give their (standard) proofs for completeness.

THEOREM 2.12. *Let $f \in L^2(\Omega^n, \mu^{\otimes n})$ be a degree-d polynomial with $||f||_2 = 1$. Then for any $t > e^{d/2}$,*

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

where $c := \frac{\alpha(\mu) d}{e}$.

Proof. For convenience let us use α instead of $\alpha(\mu)$ and set $p = \frac{2\alpha}{e} t^{2/d}$. By Markov's inequality, we have

$$\Pr[|f| > t] = \Pr[|f|^p > t^p] \leq \frac{||f||_p^p}{t^p}. \tag{2.1}$$

Now, by Theorem 2.7 and Fact 2.8 we have

$$||f||_p \leq C_p(\alpha)^{d/2} ||f||_2 \leq \frac{t}{e^{d/2}}.$$

Note that Theorem 2.7 is only applicable for $p \geq 2$, but in the case $p \in [1, 2]$ the bound $||f||_p \leq \frac{t}{e^{d/2}}$ follows trivially from the monotonicity of ℓ_p norms and the assumption $t > e^{d/2}$.

Plugging this into Equation (2.1) we get

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

□

THEOREM 2.13. *Let $f \in L^2(\Omega^n, \mu^{\otimes n})$ be a degree-2 polynomial with $||f||_2 = 1$, and let $x_1, \ldots, x_m$ be i.i.d. from $(\Omega^n, \mu^{\otimes n})$. Then, for every $r > 0$ satisfying $r < \frac{e}{\alpha(\mu)}\sqrt{m}$, it holds that*

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

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

Proof. Let us again use α instead of $\alpha(\mu)$. By Markov's inequality and the standard Chernoff method, we have

$$\Pr\left[\sum_{i=1}^m f(x_i) - m\,\mathbb{E}[f] > r\sqrt{m}\right] \leq \frac{\prod_{i=1}^m \mathbb{E}[\exp(\lambda f(x_i))]}{\exp(\lambda m\,\mathbb{E}[f] + \lambda r\sqrt{m})}. \tag{2.2}$$

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

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

where the second inequality used $\mathbb{E}[f(x_i)^k] \leq \left(\frac{k}{2\alpha}\right)^k$ (Theorem 2.7 and Fact 2.8) and $k! \geq (k/e)^k$. Assuming that λ is small enough so that $\frac{\lambda e}{2\alpha} < 1/2$, we then get

$$\mathbb{E}[\exp(\lambda f(x_i))] \leq 1 + \lambda\,\mathbb{E}[f] + \frac{\left(\frac{\lambda e}{2\alpha}\right)^2}{1 - \frac{\lambda e}{2\alpha}} \leq \exp\left(\lambda\,\mathbb{E}[f] + \frac{1}{2}\left(\frac{\lambda e}{\alpha}\right)^2\right).$$

Hence, the bound in Equation (2.2) becomes

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

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

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

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 degree-2 polynomial $-f$. 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. □

3. Limited Independence and Low-Degree Polynomials. First, we characterize the sets $X \subseteq \Omega^n$ which support k-wise independent distributions, in terms of degree-k polynomials over Ω^n. We begin with the following easy lemma, which is a straightforward generalization of the well-known fact that a distribution over bits has uniform distribution if and only if the exclusive OR of any non-empty subset of the bits is unbiased.

LEMMA 3.1. *Let $(\Omega^n, \mu^{\otimes n})$ be a finite product space with Fourier basis $\{\chi_\sigma\}_{\sigma \in \mathbb{Z}_q^n}$, and let (Ω^n, η) be an arbitrary probability space. Then $\eta = \mu^{\otimes n}$ if and only if*

$$\mathop{\mathbb{E}}_{x' \in (\Omega^n, \eta)}[\chi_\sigma(x')] = 0$$

for every $\sigma \in \mathbb{Z}_q^n$ with $|\sigma| > 0$.

Proof. Define $f : \Omega^n \to \mathbb{R}$ by $f(x) = \eta(x)/\mu^{\otimes n}(x)$. Note that $\eta = \mu^{\otimes n}$ iff f is a constant, i.e., iff $\mathrm{Var}[f] = 0$, which happens iff $\hat{f}(\sigma) = 0$ for every $\sigma \neq \mathbf{0}$. Let us then compute $\hat{f}$. We have

$$\begin{aligned}\hat{f}(\sigma) = \langle \chi_\sigma, f \rangle_{\mu^{\otimes n}} &= \mathop{\mathbb{E}}_{x \in (\Omega^n, \mu^{\otimes n})}[\chi_\sigma(x)\eta(x)/\mu^{\otimes n}(x)] \\ &= \sum_{x \in \Omega^n} \mu^{\otimes n}(x)\chi_\sigma(x)\eta(x)/\mu^{\otimes n}(x) = \mathop{\mathbb{E}}_{x \in (\Omega^n, \eta)}[\chi_\sigma(x)].\end{aligned}$$

Thus, $\eta = \mu^{\otimes n}$ if and only if

$$\mathop{\mathbb{E}}_{x \in (\Omega^n, \eta)}[\chi_\sigma(x)] = 0$$

for all $\sigma \neq \mathbf{0}$, as desired. □

We now state the characterization of the subsets of Ω^n that support k-wise independent distributions.

THEOREM 3.2. *Let (Ω, μ) be a finite probability space, and let $X \subseteq \Omega^n$ be a set of strings over Ω. Then, the following conditions are equivalent:*

(1) There exists a k-wise independent distribution η over Ω^n with marginals μ (i.e., $\eta_S = \mu^{\otimes |S|}$ for every $|S| \leq k$) such that $\mathrm{Supp}(\eta) \subseteq X$

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

(3) There is no degree k polynomial $f \in L^2(\Omega^n, \mu^{\otimes n})$ such that $f(x) > \mathbb{E}[f]$ for every $x \in X$.

Note that while item (2) of the above characterization does not explicitly mention the underlying space (Ω, μ) it is the case that $X^{:\leq k:}$ depends on the space through the characters. 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 Ω^n supported on X. Any convex combination $\sum_{x \in X} c_x \cdot x^{:\leq k:} \in \mathrm{Conv}(X^{:\leq k:})$ corresponds to the probability distribution η_c over Ω^n in which

$$\eta_c(x) = \begin{cases} c_x & \text{if } x \in X \\ 0 & \text{otherwise} \end{cases}.$$

Thus, it suffices to prove that, for every convex combination $\{c_x\}_{x \in X}$, the corresponding distribution η_c has all k-dimensional marginals being the uniform distribution iff $\sum c_x \cdot x^{:\leq k:} = \mathbf{0}$. This in turn follows from Lemma 3.1.

(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 $\left\langle v, x^{:\leq k:} \right\rangle > 0$ 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. *For every probability space (Ω, μ) there is a constant $c := \alpha(\mu)/25$ such that for any degree-d polynomial $f \in L^2(\Omega^n, \mu^{\otimes n})$ with $\mathbb{E}[f] = 0$ and $\mathrm{Var}[f] = 1$, it holds that*

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

A similar statement can be found in [8] for the Boolean case, $\Omega = \{-1, 1\}$. 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 Corollary 2.10. Define $g \in L^2(\Omega^n, \mu^{\otimes n})$ 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 lower bound $\Pr[f > c^d] = \Pr[g > 0]$ by the second moment method:

$$\Pr[g > 0] \geq \frac{\mathbb{E}[g]^2}{\mathbb{E}[g^2]} > ||g||_1^2, \tag{4.1}$$

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 Corollary 2.10, is lower-bounded by

$$\begin{aligned} ||g||_1 &\geq \frac{1}{2}\left(\frac{\alpha(\mu)}{4}\right)^{d/2} ||f||_2 - c^d = \left(\frac{\alpha(\mu)}{4}\right)^{d/2} \left(\frac{1}{2} - \frac{\alpha(\mu)^{d/2}}{(25/2)^d}\right) \\ &\geq \left(\frac{\alpha(\mu)}{4}\right)^{d/2} \cdot \frac{1}{(25/4)^{d/2}} = c^{d/2} \end{aligned}$$

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 $X \subseteq \Omega^n$ of sufficiently large constant density supports a k-wise independent distribution.

COROLLARY 4.2. *Let (Ω, μ) be a finite probability space. Then every set $X \subseteq \Omega^n$ of density $\mu^{\otimes n}(X) \geq 1 - (\alpha(\mu)/4)^k/4$ supports a k-wise independent distribution with marginals μ.*

Proof. This is almost a direct consequence of Theorem 3.2 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

$$||g||_1 = \frac{1}{2}||f||_1 \geq \frac{1}{2}\left(\frac{\alpha(\mu)}{4}\right)^{k/2} ||f||_2$$

and the corollary follows from (4.1). □

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 the Boolean hypercube equipped with the uniform distribution, and 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 $1 - 2^{-d}$.

4.1. The Boolean Hypercube. Because of our application to approximation resistance of predicates, the case of pairwise independence in the Boolean hypercube with the uniform distribution is of special interest to us, and we now examine how Corollary 4.2 can be improved in this setting.

The bound in Corollary 4.2 is based on the relation between the ℓ_2 norm and the ℓ_1 norm. Using Theorems 2.11 and 2.9 one gets the bound $||f||_2 \leq 15^{1/2}||f||_1$ for degree-2 polynomials in $L^2(\{-1,1\}^n,\mathcal{U})$. This in turn improves the bound for $k = 2$ in Corollary 4.2 from $255/256$ to $59/60$. As an alternative approach Ryan O'Donnell has suggested the following proof along the lines of the proof [8] for their variant of Theorem 4.1, giving an even better bound of $32/33$.

THEOREM 4.3. *Let $f \in L^2(\{-1,1\}^n,\mathcal{U})$ 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 2.11 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 [4] we get the following result.

COROLLARY 4.4. *Let P be any predicate on t Boolean variables 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 2.11, 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.

Finally, we give a lower bound on the density needed for a subset $X \subseteq \{-1,1\}^t$ to be certain to support a pairwise independent distribution. We have the following theorem, saying that given a subset of the hypercube that does not support pairwise independence, there is a strictly denser subset of a hypercube in higher dimension that also does not support pairwise independence.

THEOREM 4.5. *Let $X \subset \{-1,1\}^t$ be such that there is no balanced pairwise independent distribution supported on X. Then there is an $t' \geq t$ and $X' \subseteq \{-1,1\}^{t'}$ with*

$|X'|/2^{t'} = |X|/2^t + 1/2^{t'}$ *such that there is no balanced pairwise independent distribution supported on* X'.

Proof. We construct X' as follows. For each $x \in X$ and $y \in \{-1,1\}^{t'-t}$, concatenate x and y and add the resulting string to X'. Finally, let x be some string *not* in X, and y some string in $\{-1,1\}^{t'-t}$, and add the concatenation of x and y to X'. Let $\tilde{x} \in \{-1,1\}^{t'}$ denote this last string added to X'.

Consider an arbitrary distribution η on X', and write it as $\eta = \delta \cdot \tilde{\eta} + (1-\delta) \cdot \eta'$, where $\tilde{\eta}$ is a point distribution on $\tilde{x}$ and η' is some distribution on $X' \setminus \tilde{x}$.

First, we have the following claim, bounding the min-entropy of pairwise independent distributions.

CLAIM 4.6. *Let η be a balanced pairwise independent distribution over $\{-1,1\}^t$. Then for every $x \in \{-1,1\}^t$ it holds that $\eta(x) \leq 1/(t+1)$.*

Proof. Fix an arbitrary $x^* \in \{-1,1\}^t$. For notational convenience, we define for $x \in \{-1,1\}^t$ an additional coordinate x_0 which is always taken to be 1. For $0 \leq i \leq t$, consider the vector $w_i \in \mathbb{R}^{2^t}$ where, for $x \in \{-1,1\}^t$, we set $w_i(x) = \sqrt{\eta(x)}x_i$, where we identify the coordinates of $\mathbb{R}^{2^t}$ with the points of $\{-1,1\}^t$ arbitrarily.

Then $||w_i||_2^2 = 1$, and by the pairwise independence of η it holds that the w_i's are orthogonal, so $||\sum_{i=0}^t x_i^* w_i||_2 = \sqrt{t+1}$. On the other hand, the x^*'th coordinate of $\sum_{i=0}^t x_i^* w_i$ equals $(t+1)\sqrt{\eta(x^*)}$ and hence $\eta(x^*) \leq 1/(t+1)$. As x^* was arbitrary the claim follows. □

Next, note that there exists some $\delta^* > 0$ (depending on X) such that if $\delta < \delta^*$ then η is not pairwise independent. To see this, define the distance of η from being pairwise independent as

$$d(\eta) = \max_{1 \leq |S| \leq 2} \left| \mathop{\mathbb{E}}_{\eta} \left[\prod_{i \in S} x_i \right] \right|, \tag{4.2}$$

and similarly the distance of X from being able to support pairwise independence as $d(X) = \inf_{\mathrm{Supp}(\eta) \subseteq X} d(\eta)$. As the set of measures on a finite set is a compact space we know that whenever X does not support a pairwise independent measure then $d(X)$ is strictly positive. Now take the measure η' and convert it into a measure, η'' on X by setting

$$\eta''(x) = \sum_{y \in \{-1,1\}^{t'-t}} \eta'(x||y)$$

where $x||y$ is the concatenation of x and y. Let $S_0 \subseteq [t]$ be the set giving the maximum value, in (4.2) for the measure η''. Clearly

$$\mathop{\mathbb{E}}_{\eta'} \left[\prod_{i \in S_0} x_i \right] = \mathop{\mathbb{E}}_{\eta''} \left[\prod_{i \in S_0} x_i \right]$$

and hence

$$d(\eta) \geq (1-\delta) d(\eta'') - \delta \geq (1-\delta) d(X) - \delta$$

so that if $\delta < \delta^* := \frac{d(X)}{1+d(X)}$ then η is not pairwise independent.

But by the claim we also see that if $\delta > 1/(t'+1)$ then η is not pairwise independent. Hence if $t'+1 > 1/\delta^*$, we conclude that η, which was an arbitrary distribution on X', can

not be balanced pairwise independent. □

PROPOSITION 4.7. *The following subset of* $\{-1,1\}^4$ *of size* 13 *can not support pairwise independence:*

$$\{\,(-1,-1,-1,-1),\,(-1,-1,-1,+1),\,(-1,-1,+1,-1),\,(-1,-1,+1,+1),$$
$$(-1,+1,-1,-1),\,(-1,+1,-1,+1),\,(-1,+1,+1,-1),\,(-1,+1,+1,+1),$$
$$(+1,-1,-1,-1),\,(+1,-1,-1,+1),\,(+1,-1,+1,-1),\,(+1,-1,+1,+1),$$
$$(+1,+1,+1,+1)\,\}$$

Proof. Note that this set obtained by applying the construction of Theorem 4.5 to the set $\{(-1,-1),(-1,+1),(+1,-1)\}$. In this case it is very simple to obtain an explicit bound on δ^*: any pairwise independent distribution must put weight $1/4$ on $(+1,+1,+1,+1)$ since this is the only string where the first two bits are $+1$. On the other hand, by Claim 4.6, a pairwise independent distribution can put weight at most $1/5$ on $(+1,+1,+1,+1)$. □

As an immediate corollary to Theorem 4.5 and Proposition 4.7, we have

COROLLARY 4.8. *There is a constant* $\delta > 0$ *and a set* $X \subseteq \{-1,1\}^n$ *such that* $|X| = (13/16+\delta)2^n$ *and* X *does not support pairwise independence.*

5. Obtaining k-wise Independence. In this section, we give an upper bound of the form $c_{q,k}n^k\log(n)$ on the threshold for randomly supported independence. This comes relatively close to matching our lower bound, established in Section 7, of $c'n^k$ 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. *For every* (Ω,μ) *there are constants* $c,\delta > 0$ *such that the following holds. Let* $x_1,\ldots,x_m \in \Omega^n$ *be a sequence of* m *independent samples from* $(\Omega^n,\mu^{\otimes n})$. *Then, if* $m > (cn)^k\log(n^k)$, *the probability that* $X = \{x_1,\ldots,x_m\}$ *contains a pairwise independent distribution with marginals* μ *is at least* $1-\exp(-\delta n^k)$

Proof. By Theorem 3.2, $x_1,\ldots,x_m$ *does not* support a k-wise independent distribution if and only if there is a degree-k polynomial $f \in L^2(\Omega^n,\mu^{\otimes n})$ 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 $i\in[m]$ is at most $(1-\tau^k)^m \le \exp(-\tau^k m)$, where $\tau=\alpha(\mu)/25$ 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 $f\in L^2(\Omega^n,\mu^{\otimes n})$ such that $\mathbb{E}[f]=0$, $\mathrm{Var}[f]\le 2$ and every Fourier coefficient of f is an integer multiple of ϵ.

We then have that $|\mathcal{F}_\epsilon| \le (1/\epsilon)^{O(d_k)} \le \exp(c_1(qn)^k\log 1/\epsilon)$ (recall the definition of d_k from Section 2.1) 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

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

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

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 $||f-\tilde{f}||_\infty \le \tau^k$ for every degree-k polynomial f, we would be done, since the existence of $f\in L^2(\Omega^n,\mu^{\otimes n})$ 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 $||f||_2 = 1$,*

$$||f - \tilde{f}||_\infty \leq \epsilon \left(\frac{qn}{\sqrt{\alpha(\mu)}} \right)^k,$$

provided this quantity is smaller than 1.

Proof. Let f' be the result of rounding every Fourier coefficient of f to its nearest multiple of ϵ. Then, for any $x \in \Omega^n$,

$$|f(x) - f'(x)| = \left| \sum_{\sigma \in D_k} (\hat{f}(\sigma) - \hat{f}'(\sigma)) \chi_\sigma(x) \right| \leq \epsilon \sum_{\sigma \in D_k} ||\chi_\sigma||_\infty \leq \epsilon \left(\frac{qn}{\sqrt{\alpha(\mu)}} \right)^k,$$

where the last step used Fact 2.5 and $|D_k| \leq (qn)^k$. 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

$$\epsilon = \left(\frac{\sqrt{\alpha(\mu)}\tau}{qn} \right)^k,$$

giving the bound

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

for c depending only on $\alpha(\mu)$, q and τ, which in turn depend only on (Ω, μ). □

6. Pairwise Independence. In this section, we give our main theorem.

THEOREM 6.1. *For every (Ω, μ) there are constants $c, \delta > 0$ such that the following holds. Let $x_1, \ldots, x_m \in \Omega^n$ be a sequence of m independent samples from $(\Omega^n, \mu^{\otimes n})$. Then, if $m > cn^2$, the probability that $X = \{x_1, \ldots, x_m\}$ contains a pairwise independent distribution with marginals μ is at least $1 - \exp(-\delta\sqrt{n})$*

We get an immediate corollary.

COROLLARY 6.2. *There are universal 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 3.2, 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 $\sqrt{m \cdot d_2}$ and we want to prove that K contains any vector of this length. For a unit vector v let the width of K in direction v be the maximal multiple of v in result when K is projected on the one-dimensional space v. Equivalently this

is maximal value of (v, x) when x ranges over K. It is easy to see that a convex set contains the ball of radius R iff the width in any direction is at least R.

Now, K consists of all $[-1, 1]$-valued linear combinations of $x_1^{:\leq 2:}, \ldots, x_m^{:\leq 2:}$ and as a consequence of hypercontractivity it turns 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 2.13 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 2.13 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.

Finally, let us remark on how the constant c in Theorem 6.1 depends on the underlying space (Ω, μ). Tracing through the proof, it is not hard to see that one can take c polynomial in $\alpha(\mu)$. Keeping careful track of the exponents, our proof gives that c can be of order $O(\frac{q^2 \log 1/\alpha(\mu)}{\alpha(\mu)^4})$. The main bottleneck in the current proof, giving rise to the $\alpha(\mu)^4$ factor, turns out to be an application of Theorem 2.13 (in Lemma 6.5 below). By being more careful in the proof of Theorem 2.13 and using the exact value of the degree 2 term in the Taylor expansion of $\exp(f)$ one can obtain a somewhat stronger version of Theorem 2.13 which in turn allows one to improve the $\alpha(\mu)^4$ factor to $\alpha(\mu)^3$.

Proof. [Proof of Theorem 6.1] Let $m > c_0 d_2$, where c_0 is a constant that is chosen sufficiently large. We 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 3.2 this implies that X contains a pairwise independent distribution. This then implies Theorem 6.1 with $c := c_0 q^2$, since $d_2 \leq q^2 n^2$.

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. To see this, note that if $\overline{x}$ that is in the interior of K it can be written as $\sum_i a_i x_i^{:\leq 2:}$ with all $|a_i| < 1$ (since the point $(1+\delta)\overline{x}$ has to be in K for some $\delta > 0$). In particular not all the a_i's are equal to 1 and therefore 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 prove that, with high probability, the minimum width of K is larger than $||\overline{x}||$ (where $||\cdot||$ denotes the standard Euclidean ℓ_2 norm in $\mathbb{R}^{d_k}$). In particular, we have the following two lemmas (where the constants involved depend solely on the underlying space (Ω, μ))

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

$$\inf_v \mathrm{Width}(K, v) < c_2 m \tag{6.1}$$

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*

$$||\overline{x}|| > 2\sqrt{m d_2} \tag{6.2}$$

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 $c_0 = \max(c_1, (2/c_2)^2)$, and $m > c_0 d_2$. Then by a union bound there is a δ such that with probability at least $1 - \exp(-\delta\sqrt{n})$, neither Equation (6.1) nor Equation (6.2) holds, and we have

$$\inf_v \mathrm{Width}(K, v) \geq c_2 m > 2\sqrt{m d_2} \geq ||\overline{x}||.$$

This implies that $\overline{x}$ lies strictly inside K, as desired. Hence, if $m > cn^2 \geq c_0 d_2$, the probability that $\mathbf{0} \in \mathrm{Conv}(X^{:\leq 2:})$ 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. [Proof of Lemma 6.4] Let

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

be the squared length of $\overline{x}$. We can then view l as a degree 4 polynomial in $L^2(\Omega^{nm}, \mu^{\otimes mn})$. Our goal is to apply the concentration bound Theorem 2.12 to l. To be successful in this, we need that the variance $\mathrm{Var}[l]$ 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

$$\begin{aligned} l^2 &= \sum_{\sigma_1, \sigma_2} \left(\sum_{i=1}^{m} \chi_{\sigma_1}(x_i) \right)^2 \left(\sum_{i=1}^{m} \chi_{\sigma_2}(x_i) \right)^2 \\ &= \sum_{\sigma_1, \sigma_2} \sum_{i_1, i_2, i_3, i_4 \in [m]} \chi_{\sigma_1}(x_{i_1}) \chi_{\sigma_1}(x_{i_2}) \chi_{\sigma_2}(x_{i_3}) \chi_{\sigma_2}(x_{i_4}). \end{aligned}$$

Define

$$S(\sigma_1, \sigma_2) = \sum_{i_1, i_2, i_3, i_4 \in [m]} \chi_{\sigma_1}(x_{i_1}) \chi_{\sigma_1}(x_{i_2}) \chi_{\sigma_2}(x_{i_3}) \chi_{\sigma_2}(x_{i_4}),$$

and let us analyze $\mathbb{E}[S(\sigma_1, \sigma_2)]$. If $\sigma_1 \neq \sigma_2$, the expected value of

$$\chi_{\sigma_1}(x_{i_1}) \chi_{\sigma_1}(x_{i_2}) \chi_{\sigma_2}(x_{i_3}) \chi_{\sigma_2}(x_{i_4})$$

is 0 unless $i_2 = i_1$ and $i_4 = i_3$. Hence for $\sigma_1 \neq \sigma_2$, we have

$$\mathbb{E}[S(\sigma_1, \sigma_2)] = \sum_{i_1, i_3} \mathbb{E}[\chi_{\sigma_1}(x_{i_1})^2 \chi_{\sigma_2}(x_{i_3})^2].$$

The terms where $i_1 \neq i_3$ contribute 1 to this sum, and the terms where $i_1 = i_3$ contribute at most $1/\alpha^2$, where $\alpha = \alpha(\mu)$, by Fact 2.5. Hence we have for $\sigma_1 \neq \sigma_2$

$$\mathbb{E}[S(\sigma_1, \sigma_2)] \leq m^2 + m/\alpha^2.$$

Now let $\sigma_1 = \sigma_2 := \sigma$, and consider the expected value of

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

If for any $j \in [m]$ it is the case that only one of the i_k:s equal j, this expectation is 0. 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 $i = j$ and $k = l$, or $i = k$ and $j = l$, or $i = l$ and $j = k$. There are exactly $3m(m-1) + m \leq 3m^2$ ways to choose i_1, i_2, i_3, i_4 in such a paired way and hence in this case

$$\mathbb{E}[S(\sigma, \sigma)] \leq 3m^2/\alpha^2,$$

where we again used Fact 2.5. After these lengthy computations we thus find that

$$\mathbb{E}[l^2] = \sum_{\sigma_1, \sigma_2} \mathbb{E}[S(\sigma_1, \sigma_2)] \leq d_2^2 m^2 + d_2^2 m/\alpha^2 + 3d_2 m^2/\alpha^2,$$

so that

$$\mathrm{Var}[l] \leq d_2^2 m/\alpha^2 + 3d_2 m^2/\alpha^2 \leq 4d_2 m^2/\alpha^2,$$

where the last inequality assumed that $m \geq d_2$. Applying Theorem 2.12 to the polynomial $(l - \mathbb{E}[l])/\sqrt{\mathrm{Var}[l]}$, we have

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

for $c' = c\sqrt{3\alpha/2}$. Since $d_2 \geq q^2 n^2$, the lemma follows with $\delta_2 = c'\sqrt{q}$. □

We now move on to the proof of Lemma 6.3. By a standard argument the width 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 $c_3, \tau > 0$ such that the following holds: for every $v \in \mathbb{R}^{d_2}$ with $||v|| = 1$, the probability that*

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

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

Proof. Set $2c_3 = \alpha(\mu)/4$, the constant from Corollary 2.10 for $d = 2$. For $v \in \mathbb{R}^{d_2}$ with $||v|| = 1$, let $f_v \in L^2(\Omega^n, \mu^{\otimes n})$ be the corresponding degree-2 polynomial such that $f_v(x) = \left\langle v, x^{:\leq 2:} \right\rangle$.

By definition,

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

The maximum is clearly attained by setting

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

so that

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

Applying Theorem 2.13 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 $\tau > 0$. 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 Corollary 2.10 lower bounded by $2c_3||f_v||_2 \cdot m = 2c_3 m$.

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

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

□

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

Proof. [Proof of Lemma 6.3] Let $V = \{v_1, \ldots, v_L\}$ be an ϵ-net of the unit sphere in $\mathbb{R}^{d_2}$ (where we are eventually going to choose $\sqrt{\epsilon}$ to be a sufficiently small multiple of c_3 from Lemma 6.5)), 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

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

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 at most

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

provided $m \geq 2d_2 \ln(1/\epsilon)/\tau$.

Let $W_{\max} = \sup_{||v||=1} \operatorname{Width}(K, v)$. 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 $u \in K$

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

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 a small constant multiple of c_3^2.

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 $u \in K$,

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

again provided ϵ is a small constant multiple of c_3^2.

Hence, with probability at least $1-\exp(-\delta m)$ (where $\delta = \tau/2$), we have $\inf_{||v||=1}\mathrm{Width}(K, v) \geq c_3 m/2 := c_2 m$, provided that $m \geq c_1 d_2$ where $c_1 = 2\ln(1/\epsilon)/\tau$. □

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. *Let Ω be a space of size $q = |\Omega|$ and $\mathcal{U}$ denote the uniform distribution over Ω. Let $x_1, \ldots, x_m$ be a sequence of m independent samples from $(\Omega^n, \mathcal{U}^{\otimes n})$. Then, if $m < \left(\frac{n}{2k^2q^k}\right)^k$, the probability that $x_1, \ldots, x_m$ can support a k-wise independent distribution with marginals $\mathcal{U}$ (i.e., a balanced k-wise independent distribution) is at most $\exp\left(-\frac{n}{4kq^k}\right)$.*

Proof. We prove that, if $m \leq \left(\frac{n}{2k^2q^k}\right)^k$, 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 $x_1^{:\leq k:}, \ldots, x_m^{:\leq k:}$ is non-zero, so that, by Theorem 3.2, $x_1, \ldots, x_m$ does not support a k-wise independent distribution.

The main component of the proof is the following lemma.

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

Before proving the lemma we finish the proof of the theorem. Set $m_0 = \frac{n}{2k^2q^k}$ and $m = m_0^k$, and let $x_1, \ldots, x_m$ be m uniformly random points of Ω^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

$$1 - m\exp\left(-\frac{n}{2kq^k}\right) = 1 - \exp\left(-k\left(m_0 - \ln(m_0)\right)\right) \geq 1 - \exp\left(-km_0/2\right).$$

This concludes the proof of Theorem 7.1. □

Next, we turn to the proof of the lemma.

Proof. [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 $x \in \Omega^n$, we have $x^{:\leq k:} \in S$ iff $f_{v_i}(x) = 0$ for every i, where we again interpret v_i as giving the coefficients of a degree-k polynomial. We prove that only an exponentially small fraction of all points $x \in \Omega^n$ satisfy these conditions.

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

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

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

$$T(n,m) \leq (q^k-1)^{n/k} \cdot \exp(km^{1/k}). \tag{7.1}$$

If $d_k(n) \leq m$ so that $n \leq m^{1/k}k/(q-1)$, we have the trivial bound $T(n,m) \leq q^n \leq \exp(km^{1/k})$, so let $d_k(n) > m$ and assume inductively that Equation (7.1) holds for all $n' < n$. Assume that there is a f_{v_i} which has degree exactly k (if all f_{v_i} have degree at most $k-1$, we would get an even better bound). Without loss of generality, we can take f_{v_1} to have degree exactly k, and having a non-zero coefficient σ with support $S(\sigma) = [k]$.

Next, eliminate (by standard Gaussian elimination) all coordinates σ' with $S(\sigma') \cap [k] \neq \emptyset$. As there are exactly $d_k(n) - d_k(n-k)$ such values of σ', 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 $x^* \in \Omega^{n-k}$, consider the number of ways of extending it to a solution for the original system. Plugging in x^* in the equation $f_{v_1}(x) = 0$, this equation becomes an equation of the form

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

for some function $p : \Omega^k \to \mathbb{R}$. Furthermore, the function p is not identically zero, since $\hat{p}(\sigma) = \hat{f}_{v_1}(\sigma) \neq 0$. This implies that the number of ways of extending x^* is at most $q^k - 1$, and hence we have

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

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

$$(q^k-1)^{n/k} \exp(km^{1/k})/q^n = (1-q^{-k})^{n/k} \exp(km^{1/k}) \leq \exp\left(-\frac{n}{kq^k} + km^{1/k}\right).$$

Plugging in $m \leq \left(\frac{n}{2k^2q^k}\right)^k$, the lemma follows. □

8. Approximating a Random Predicate. In this section we let P be a predicate constructed by randomly choosing $O(t^2/\log t)$ strings of length t and making these be the inputs accepted by P. We then show that with high probability this predicate is not approximation resistant. Formally:

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

8.1. The Boolean Case. As the case of general domains gives a more complicated argument we begin by establishing the theorem in the case of the Boolean domain which illustrates the main idea.

THEOREM 8.2. *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 [12].

LEMMA 8.3. *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 8.4. *Suppose P is constructed as in the hypothesis of Theorem 8.2, 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 8.3 and Lemma 8.4 jointly imply Theorem 8.2 and thus all we need to do is to establish Lemma 8.4.

Proof. [Proof of Lemma 8.4] $P^{=2}$ 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

$$\begin{aligned} P^{=2}(y) &= 2^{-t} \sum_{i<j} \sum_{z \in P^{-1}(1)} z_i z_j y_i y_j. \\ &= 2^{-t} \left(\binom{t}{2} + \sum_{\substack{z \in P^{-1}(1) \\ z \neq y}} \sum_{i<j} z_i z_j y_i y_j \right) \end{aligned} \tag{8.1}$$

The sum in Equation (8.1) 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 2.13, 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

$$\Pr\left[\left| \sum_z P_y(z) \right| \geq \binom{t}{2} \right] \leq \exp(-\Omega(t^2/s)) \leq 1/t^3 \tag{8.2}$$

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

8.2. Arbitrary Domains. Let us then turn to the proof in the case of general size domains. The proof follows the ideas of [12] and [14] and is completely analogous to the Boolean case, but as the relevant analogue of Lemma 8.3 does not seem to have appeared we need supply some details of this part of the argument even though the extension is immediate.

First we need to define literal and, as done in [4] we let it be of the form $x + a$ where x is a variable, a is a constant and addition is done modulo q. [1]

Let us start be giving an overview of the approach. The idea is to construct numbers $p_i^a, i \in [n], 0 \leq a \leq q-1$ such that $-1 \leq p_i^a \leq q-1$ and

$$\sum_{a=0}^{q-1} p_i^a = 0 \tag{8.3}$$

and then set x_i to a with probability $(1+\epsilon p_i^a)/q$ for a small number $0 < \epsilon \leq 1$, independently for each i.

If we have a total of m constraints each satisfied by s inputs then it easy to see that the expected number of satisfied constraints is of the form

$$\frac{ms}{q^t} + \frac{\epsilon L(\mathbf{p}) + \epsilon^2 Q(\mathbf{p}) + \epsilon^3 P'(\mathbf{p}, \epsilon)}{q^t}, \tag{8.4}$$

where $\mathbf{p}$ is the vector of p_i^a, L is a linear function, Q is a bilinear function, and $\epsilon^3 P'$ contains all terms that are of degree at least 3. Let us analyze this expression a little bit more carefully.

We have

$$P(x) = \sum_{z \in P^{-1}(1)} P^z(x) \tag{8.5}$$

where $P^z(x)$ is the predicate which is true if and only if $x = z$ and let us first try to prove that there is some set of probabilities that makes $Q(p)$ large. Let us first analyze a single application of the predicate. For an assignment $y \in [q]^t$ let us define a set of probabilities by setting $p_i^a = q-1$ when $y_i = a$ and $p_i^a = -1$ otherwise. The probability that a random assignment, with this probability distribution, satisfies P is of the form

$$\frac{s}{q^t} + \frac{\epsilon L^y + \epsilon^2 Q^y + \epsilon^3 P^y(\epsilon)}{q^t}.$$

We have the following lemma that takes the place of Lemma 8.4.

LEMMA 8.5. *Suppose P is constructed as in the hypothesis of Theorem 8.1, then for any $y \in P^{-1}(1)$,*

$$\Pr[Q^y > 0] \geq 1 - t^{-3}.$$

Proof. Using the expansion (8.5) we have contributions to Q^y from $z = y$ and from other terms. The contribution from y is $\binom{t}{2}(q-1)^2$ and let us analyze the contribution from a randomly chosen z. This contribution is the value of a quadratic polynomial P_y evaluated at a random point and it has expectation zero. Indeed, each variable of P_y indicate whether the randomly chosen z equals y in that particular coordinate. If the coordinate is equal,

[1] In fact, the current proof works in the more general context that a literal is a variable together with a permutation π mapping the value of the variable to a value of the literal.

which happens with probability $1/q$ then the variable takes the value $q-1$ and otherwise it takes the value -1. Thus $P_y(z)$ is the sum $\binom{t}{2}$ terms which each has the value $(q-1)^2$ with probability $1/q^2$, the value $-(q-1)$ with probability $2(q-1)/q^2$ and the value 1 with probability $(q-1)^2/q^2$.

It is not difficult to see that each value $P_y(z)$ has variance $\binom{t}{2}(q-1)^2$ and bounding the probability that $\sum_z P_y(z)$ is negative can be done in way analogous to the Boolean case. We omit the details. □

Next we prove that from this local property we get that the overall quadratic part, $Q(p)$ in (8.4), is large.

LEMMA 8.6. *There is a constant $c_q > 0$ such that the following is true. Suppose $s \leq c_q t^2/\log t$ and suppose $P : [q]^t \mapsto \{0,1\}$ is a predicate chosen randomly among all predicates that accept s inputs. Then, with probability $1 - o_t(1)$ there exist $\epsilon' > 0$ and $\delta > 0$ such that whenever the number of simultaneous satisfiable constraints is at least $(1-\epsilon')m$ there is a choice of values p_i^a such that $Q(\mathbf{p}) > \delta m$.*

Proof. With probability $1 - o(t)$, Lemma 8.5 is true for all $y \in P^{-1}(1)$ and define

$$\epsilon_0 = \min_{y \in P^{-1}(1)} Q^y.$$

Now, given the assignment x that satisfies $(1-\epsilon')m$ clauses we set $p_i^a = q-1$ if $x_i = a$ and $p_i^a = -1$ otherwise. The contribution to Q of any satisfied constraint is at least ϵ_0, while the contribution to Q for any non-satisfied constraint is lower bounded by a constant c which can be taken to be $-s\binom{t}{2}(q-1)^2$. Thus

$$Q(p) \geq (1-\epsilon')m\epsilon_0 - \epsilon' mc$$

and choosing ϵ' to be a sufficiently small constant this is lower bounded by δm for $\delta > 0$. □

Next we prove that if we can make $Q(p)$ large then we can indeed find a good assignment.

LEMMA 8.7. *For any $\delta > 0$ there is a $c_\delta > 0$ such that if $Q(\mathbf{p}) \geq \delta m$ then there is a value of ϵ such that setting $x_i = a$ with probability $\frac{1+\epsilon p_i^a}{q}$ satisfies $m(s + c_\delta)q^{-t}$ constraints in expectation.*

Proof. Let us first note that, provided we choose $\epsilon \leq \frac{1}{q-1}$ to maintain non-negative probabilities, we may change p_i^a to $-p_i^a$ and hence we can assume that $L(\mathbf{p}) \geq 0$. In order to lower bound (8.4) we need to upper bound $|P'(\mathbf{p}, \epsilon)|$, which we do in the Claim 8.8 below.

Now, setting $\epsilon = \delta(2s(2(q-1))^t)^{-1}$ (which is upper-bounded by $\frac{1}{q-1}$ as needed) we see, by the claim below that $|\epsilon^3 P'(\mathbf{p}, \epsilon)| \leq \epsilon^2 Q(\mathbf{p})/2$ and thus we can choose $c_\delta = \epsilon^2\delta/2$. This concludes the proof of Lemma 8.7. □

CLAIM 8.8. *For every $0 \leq \epsilon \leq 1$, $|P'(\mathbf{p}, \epsilon)| \leq ms(2(q-1))^t$.*

Proof. We use the expansion (8.5). The expected value of each $P^z(x)$ is exactly the probability that x takes the value z. This probability is of the form $\prod_{i=1}^t \frac{1+\epsilon p_i^{z_i}}{q}$: expanding that product we get at most 2^t terms and using $|p_i^a| \leq q-1$ and $\epsilon \leq 1$ we get a bound $(2(q-1))^t$ for the contribution of a single occurrence of P^z to P'. We have m constraints and each is the sum of s different delta-predicates. The claim follows. □

Finally, we can find reasonable approximations to the maximum value of $Q(p)$.

LEMMA 8.9. *There is a constant factor approximation algorithm for $Q(p)$.*

Proof. [Sketch of proof] Though we proved existence of numbers $\mathbf{p}$ making $Q(\mathbf{p})$ large we also need to discuss how to find such values efficiently. This is done, as is standard, by semidefinite programming.

We replace each variable p_i^a by a vector v_i^a and replacing products by scalar products we can maximize $Q(\mathbf{v})$. The obtained vector solution can be translated back to biases using exactly the same procedure as used in [14].

Namely we pick a random vector r where each coordinate is normally distributed with mean zero and standard deviation 1 and set $y_i^a = (r, v_i^a)$.

This ensures that $E_r[Q(\mathbf{y})] = Q(\mathbf{v})$ and the only problem is that we have a small probability of y_i^a being too large in absolute value. The remedy is to use $\frac{1}{D} y_i^a$ whenever $|y_i^a| \leq D$ and setting $y_i = 0$ otherwise.

This decreases the value of $Q(\mathbf{y})$ by a factor of D^2, but as the probability of y_i^a being too large decreases as $e^{-D^2/2}$ the loss from large values of y_i^a is insignificant for sufficiently large D. We omit the details as the argument follows very close the argument of [14]. □

Theorem 8.1 now follows from the Lemmas 8.6, 8.7 and 8.9.

We remark that using this approach one can also prove that there is a constant $c(q)$ such that *every* predicate $P : [q]^t \to \{0,1\}$ with at most $c(q) \cdot t$ accepting assignments is approximable, analogously to the theorem by Hast [12] that any predicate $P : \{-1,1\}^t \to \{0,1\}$ on t Boolean inputs having fewer than $2\lceil (t+1)/2 \rceil$ accepting inputs is approximable. In this case one changes Lemma 8.6 by letting P be an arbitrary predicate (as opposed to a random one) with at most $c(q)t$ accepting assignments (as opposed to $c(q)t^2/\log t$ accepting assignments). We omit the details.

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 two anonymous referees for help in improving the quality of the manuscript.

REFERENCES

[1] S. Arora, B. Barak, and D. Steurer. Subexponential algorithms for unique games and related problems. To appear at FOCS 2010.

[2] 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.

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

[4] P. Austrin and E. Mossel. Approximation Resistant Predicates from Pairwise Independence. *Computational Complexity*, 18(2):249–271, 2009.

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

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

[7] 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.

[8] I. Dinur, E. Friedgut, G. Kindler, and R. O'Donnell. On the Fourier tails of bounded functions over the discrete cube. *Israel Journal of Mathamatics*, 160:389–412, 2007.

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

[10] 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.

[11] 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.

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

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

[14] J. Håstad. Every 2-CSP Allows Nontrivial Approximation. *Computational Complexity*, 17:549–566, 2008.

[15] J. Håstad. On the approximation resistance of a random predicate. *Computational Complexity*, 18:413–434, 2009.

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

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

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

[19] A. Kolla. Spectral Algorithms for Unique Games. In *IEEE Conference on Computational Complexity (CCC)*, pages 122–130, 2010.

[20] E. Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19:1713–1756, 2010.

[21] 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.

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

[23] L. Trevisan. Approximation Algorithms for Unique Games. *Theory of Computing*, 4(1):111–128, 2008.

[24] P. Wolff. Hypercontractivity of Simple Random Variables. *Studia Math.*, 180:219–236, 2007.

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