2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)

Towards a Proof of the Fourier–Entropy Conjecture?

Esty Kelman
School of Computer Science
Tel Aviv University
Tel Aviv, Israel
estykelman@gmail.com

Guy Kindler
Einstein Institute of Mathematics
Hebrew University of Jerusalem
Jerusalem, Israel
wgkindler@gmail.com

Noam Lifshitz
Einstein Institute of Mathematics
Hebrew University of Jerusalem
Jerusalem, Israel
noamlifshitz@gmail.com

Dor Minzer
Department of Mathematics
MIT
Cambridge, MA, USA
minzer.dor@gmail.com

Muli Safra
School of Computer Science
Tel Aviv University
Tel Aviv, Israel
muli.safra@gmail.com

***Abstract*—The total influence of a function is a central notion in analysis of Boolean functions, and characterizing functions that have small total influence is one of the most fundamental questions associated with it. The KKL theorem and the Friedgut junta theorem give a strong characterization of such functions whenever the bound on the total influence is $o(\log n)$. However, both results become useless when the total influence of the function is $\omega(\log n)$. The only case in which this logarithmic barrier has been broken for an interesting class of functions was proved by Bourgain and Kalai, who focused on functions that are symmetric under large enough subgroups of S_n.**

In this paper, we build and improve on the techniques of the Bourgain-Kalai paper and establish new concentration results on the Fourier spectrum of Boolean functions with small total influence. Our results include:

1) **A quantitative improvement of the Bourgain–Kalai result regarding the total influence of functions that are transitively symmetric.**
2) **A slightly weaker version of the Fourier–Entropy Conjecture of Friedgut and Kalai. Our result establishes new bounds on the Fourier entropy of a Boolean function f, as well as stronger bounds on the Fourier entropy of low-degree parts of f. In particular, it implies that the Fourier spectrum of a constant variance, Boolean function f is concentrated on $2^{O(I[f]\log I[f])}$ characters, improving an earlier result of Friedgut. Removing the $\log I[f]$ factor would essentially resolve the Fourier–Entropy Conjecture, as well as settle a conjecture of Mansour regarding the Fourier spectrum of polynomial size DNF formulas.**

Our concentration result for the Fourier spectrum of functions with small total influence also has new implications in learning theory. More specifically, we conclude that the class of functions whose total influence is at most K is agnostically learnable in time $2^{O(K\log K)}$ using membership queries. Thus, the class of functions with total influence $O(\log n/\log\log n)$ is agnostically learnable in $\mathsf{poly}(n)$ time.

***Keywords*-Fourier analysis, Fourier-Entropy Conjecture, Learning sparse functions.**

I. Introduction

The field of Analysis of Boolean functions is by now an integral part of Theoretical Computer Science, Combinatorics and Probability. Many basic results, such as the KKL Theorem [18] and the various junta theorems [10], [5] have a wide range of applications including PCP constructions [16], [23], [9], metric non-embeddability results [24], Extremal Combinatorics [8], [21] and many more.

Perhaps the most basic, non-trivial, question in the field is to characterize functions that have small total influence. Throughout the paper, we will consider the Boolean hypercube $\{0,1\}^n$ equipped with the uniform measure, [1] and in particular Boolean functions defined over it, i.e. $f\colon \{0,1\}^n \to \{0,1\}$. The influence of the ith variable, $I_i[f]$, is the probability that $f(x) \neq f(x \oplus e_i)$ when we sample x according to the uniform distribution. The total influence of f is the sum of all individual influences, i.e. $I[f] = I_1[f] + \ldots + I_n[f]$. What can be said about a function f with constant variance, that has small total influence, i.e. $I[f] \leqslant K$?

One obvious example of such functions are *K-juntas*, i.e. functions f that depend on at most K variables. Clearly, if f is a K-junta, then $I[f] \leqslant K$. A better example, known as the Tribes function, was given by Ben-Or and Linial [3], and it depends on $e^{\Omega(K)}$ variables, all of which have the same influences. The KKL Theorem and the Friedgut junta theorem state that in a sense, these are the worst possible examples. The KKL Theorem [18] asserts that in this case, there must be a variable i with a large individual influence of $e^{-O(K)}$. Friedgut [10] strengthened that result, showing that f in fact must essentially depend only on $e^{O(K)}$ variables. We note that these results are very effective when K is thought of as constant, and remain meaningful as long

[1] Many of the statements we give have natural analogs for the p-biased measure.

2575-8454/20/$31.00 ©2020 IEEE
DOI 10.1109/FOCS46700.2020.00032

as $K \leqslant \varepsilon \log n$. When K is, say, $100 \log n$, these results become completely trivial – this is a well known barrier in analysis of Boolean functions, regarding which very little is known.

Sharp thresholds: Motivated by the study of sharp thresholds of graph properties, Bourgain and Kalai [6] studied the above question for functions $f\colon \{0,1\}^n \to \{0,1\}$ that are symmetric with respect to a subgroup $G \subseteq S_n$.[2] They showed that if the subgroup G is nice enough, then one has $I[f] = \omega(\log n)$ for all functions f symmetric under G. More precisely, for each $\varepsilon > 0$ and subgroup G, Bourgain and Kalai consider the parameter $a_\varepsilon(G)$, defined to be largest d such that the orbit of each $S \subseteq [n]$ of size at most d is at least of size $e^{|S|^{1+\varepsilon}}$. Using this parameter, Bourgain and Kalai prove that $I[f] \geqslant C(\varepsilon) a_\varepsilon(G) \mathsf{var}(f)$, for some $C(\varepsilon) > 0$. The class of subgroups for which this statement is useful includes symmetries of graphs, hypergraphs, $GL(n, \mathbb{F}_q)$ and more. The latter result played an important part in a recent result in coding theory, showing that Reed-Muller codes with constant rate achieve capacity over erasure channels with random errors [27] (though by now, an alternative argument that bypasses the use of Bourgain-Kalai is known [26]).

Learning theory: Proving strong structural results on Boolean functions with bounded total influence, say at most $100 \log n$, if often times useful in designing learning algorithms for specific concept classes, such as the class of polynomial size DNF formulas. In particular, it has long been known that a concept class in which each function can be approximated by a sparse polynomial is learnable with membership queries [12], [28]. More recently, it has been shown that such property is strong enough to imply learnability in the presence of some errors, i.e. in the agnostic learning model [14]. We elaborate on the topic and on our results along these lines in Section I-C1.

Percolation: Another motivation to develop tools bypassing the logarithmic barrier comes from percolation theory. Kalai [19] asked whether there is a variant of the Bourgain-Kalai Theorem in which the symmetry condition is relaxed to a weaker notion of regularity. His question was motivated by a problem in percolation theory, in which one has a sequence of function f_n (which is the indicator of the crossing event in the 3-dimensional grid at the critical probability) and the goal is to prove good lower bounds on the total influence of f_n. More precisely, the goal is to prove that for every n one has $I[f_n] \geqslant a_n > 0$, where the sequence a_n satisfies that $\sum\limits_{n=1}^{\infty} \frac{1}{n a_n}$ converges (i.e. morally that a_n is slightly larger than $\log n$). The class of functions f_n, however, does not have the symmetries required for the Bourgain-Kalai Theorem.

Our main result can be viewed as a variant of the Bourgain-Kalai Theorem that relaxes the symmetry condition, and we prove it is enough that all low Fourier coefficients of f are small.

A. The Fourier–Entropy Conjecture

Another form of structural results on Boolean functions with $I[f] \leqslant K$ one may hope for, is that of concentration of the Fourier Spectrum only on a small number of characters. In other words, can we say that except for a negligible mass, all Fourier weight of f is concentrated on few Fourier coefficients? Friedgut's theorem [10] (or rather, its proof) implies that except for negligible amount of mass, all Fourier weight of f lies on at most $e^{O(K^2)}$ Fourier coefficients; in general, this is the best bound known to date. Friedgut and Kalai [11] conjectured that the actual answer should be $e^{O(K)}$; in fact, they propose the more refined *Fourier–Entropy Conjecture*, stating that the Shannon-Entropy of the Fourier spectrum of a Boolean function (thought of as a distribution) is at most $O(I[f])$. Here, the Fourier entropy of a function f is given by $H[\widehat{f}] = \sum\limits_{S \subseteq [n]} \widehat{f}(S)^2 \log(1/\widehat{f}(S)^2)$. Despite significant efforts, progress towards the Fourier–Entropy Conjecture has been slow, and it has been proved only for special classes of functions [7], [25], [31], [32], [34].

The min-entropy of the Fourier Spectrum of a function f is defined by $H_\infty[\widehat{f}] = \min_S \log(1/\widehat{f}(S)^2)$. The Fourier–Min–Entropy Conjecture is a relaxation of the Fourier–Entropy Conjecture, stating that min-entropy of the Fourier spectrum of a Boolean function is at most $O(I[f])$. As the min-entropy of a distribution is always upper bounded by the Shannon-Entropy of a distribution, one sees that this conjecture is strictly weaker. O'Donnell noted that for monotone functions (and more generally for unate functions), this conjecture follows immediately from the KKL Theorem (while the Fourier–Entropy Conjecture is not known for monotone functions). [3] Progress towards this conjecture has also been slow [1], [33].

B. Main results

Our main results are new bounds on the Fourier min-entropy and the Fourier entropy of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$. It will be convenient to state them in terms of the normalized total influence of f, i.e. $\tilde{I}[f] = \frac{I[f]}{\mathsf{var}(f)}$. First, we show that the Fourier min-entropy Conjecture holds up to poly-logarithmic factor in $\tilde{I}[f]$.

Theorem I.1. *There is $C > 0$, such that for any function $f\colon \{0,1\}^n \to \{0,1\}$ there is a non-empty $S \subseteq [n]$ of size*

[2] We say f is symmetric under a permutation $\pi \in S$, if $f(x) = f(\pi(x))$ for all $x \in \{0,1\}^n$, where $\pi(x)_i = x_{\pi(i)}$. We say f is symmetric under a set of permutations $G \subseteq S_n$ if it is symmetric under each $\pi \in G$.

[3] A function $f\colon \{0,1\}^n \to \{0,1\}$ is said to be increasing (respectively decreasing) with respect to coordinate i, if for every $x \in \{0,1\}^n$ with $x_i = 0$, it holds that $f(x \oplus e_i) \geqslant f(x)$ (respectively $f(x \oplus e_i) \leqslant f(x)$). The function f is called monotone if it is increasing along each $i \in [n]$, and called unate if along each $i \in [n]$ it is either increasing or decreasing.

$O(\tilde{I}[f])$ *such that* $\left|\widehat{f}(S)\right| \geqslant 2^{-C\cdot|S|\log(1+\tilde{I}[f])}\sqrt{\mathsf{var}(f)}$. *In particular,*

$$H_\infty[f] \leqslant O\left(\tilde{I}[f]\log(1+\tilde{I}[f])\right).$$

Our result is in fact stronger in several ways. First, we show that not only there exists a significant Fourier coefficient for f, but in fact almost all the Fourier mass of f lies on such characters.

Theorem I.2. *For every $\eta > 0$, there exists $C > 0$, such that for all $f\colon \{0,1\}^n \to \{0,1\}$ we have*

$$\sum_S \widehat{f}(S)^2 1_{\left|\widehat{f}(S)\right| \leqslant 2^{-C\cdot\tilde{I}[f]\log(1+\tilde{I}[f])}} \leqslant \eta \cdot \mathsf{var}(f). \quad (1)$$

Second, we show that a slightly weaker version of the Fourier–Entropy Conjecture holds for the low-degree parts of f. Here, the part of f of degree at most d is denoted by $f^{\leqslant d}$ and is defined to be the part of the Fourier transform of f up to degree d (see Section II for a formal definition).

Theorem I.3. *There exists an absolute constant $K > 0$, such that for any $D \in \mathbb{N}$ and $f\colon \{0,1\}^n \to \{0,1\}$ we have that*

$$H[\widehat{f^{\leqslant D}}] \leqslant K \sum_{|S|\leqslant D} |S|\log(|S|+1)\widehat{f}(S)^2 + K\cdot I[f].$$

Note that as $I[f] = \sum_S |S|\,\widehat{f}(S)^2$, Theorem I.3 just falls short of proving the Fourier entropy conjecture by a factor of $\log(|S|)$. In the worst case, this factor may be as large as $\log(\mathsf{deg}(f))$, however for most interesting functions the contribution of this logarithmic factor is typically much smaller. We also note that a particularly interesting setting of D is $D = 100\cdot I[f]$, since most of the Fourier mass of f lies on characters S such that $|S| \leqslant D$ and then the logarithmic factor contributes at most $\log(D)$.

The above results follow from our main technical result, Corollary V.1, which may be of independent interest. We remark that this project began with the goal of gaining a better understanding of the (notoriously hard) Bourgain-Kalai paper, and explore the underlying idea that allowed them to bypass the logarithmic barrier. Indeed, our proofs build and improve upon the ideas of Bourgain and Kalai, and in our view are also considerably simpler. The core idea of the argument now boils down to a general statement that upper bounds inner products $\langle f,g\rangle$ for a Boolean function f and a real-valued, low-degree function g, by their Fourier coefficients, total influences and norms. This result could be thought of a successive series of inequalities that improve each other, the first one of which is the KKL Theorem, and the proof of each inequality uses the previous inequalities (formally, by induction). We defer a more detailed discussion of our techniques and proofs to Section IV-A.

Remark I.4. *One can establish variants of Theorems I.1, I.2 and I.3 for general real-valued functions $f\colon \{0,1\}^n \to \mathbb{R}$. The proof goes along the same lines, except that the quantity $I[f,g]$ from Definition II.5 has to be defined differently.*

C. Applications

1) Learning Theory: The Fourier–Entropy Conjecture is closely to related to the problem of learning functions in the membership model. Intuitively, if a function has small Fourier Entropy, that that means that its Fourier transform is concentrated on a few characters. In other words, it means that the function can be approximated by a sparse polynomial, a class that is very important in the context of learning theory.

It is well-known that the class of functions approximated by sparse polynomials is learnable using membership queries [12], [28]. However, one may wonder if this class is learnable in the more challenging model of agnostic learning.

The framework of agnostic learning was defined by Kearn et al.[20], who proposed it as a more realistic version of the PAC-learning model. Indeed, it is aimed at capturing the intuition that often in the task of learning a function, the data we see is in fact a noisy version of the actual data in the real world, and therefore one cannot make very strong structural assumptions on it, but rather that it is close to a very structured object. We formalize this notion next.

A notable example in this context is the class of polynomials-size DNF formulas, which is known to be somewhat sparse (its Fourier spectrum is concentrated on at most $n^{O(\log\log n)}$ characters) but not enough

Let $f\colon \{0,1\}^n \to \{0,1\}$ be an *arbitrary* Boolean function and let $\mathcal{C} \subset \{g \mid g\colon \{0,1\}^n \to \{0,1\}\}$ be a concept class. Define $\mathsf{opt}_{\mathcal{C}}(f) = \min_{c\in\mathcal{C}} \Pr_{x\in\{0,1\}^n}[c(x) \neq f(x)]$, i.e. the distance of f from the concept class $\mathcal{C}$.

Definition I.5. *We say that $\mathcal{C}$ is* agnostically learnable *with queries with respect to the uniform distribution, if there exists a randomized algorithm that given black box access to any f, runs in time poly$(n,\varepsilon^{-1},\log\frac{1}{\delta})$ and outputs a hypothesis h such that $\Pr_{x\in\{0,1\}^n}[h(x)\neq f(x)] \leqslant \mathsf{opt}_{\mathcal{C}}(f)+\varepsilon$ with probability $1-\delta$.*

Golpalan et al.[14] showed that every concept class that has sparse approximation is agnostically learnable (the running time depends on the sparsity of the approximation and other parameters).

Theorem I.6 ([14], Theorem 16)**.** *Let $t,\varepsilon > 0$ and suppose $\mathcal{C}$ is a concept class such that for every $c \in \mathcal{C}$, there exists a $p\colon \{0,1\}^n \to \mathbb{R}$ such that*

1) *The polynomial p is t-sparse, i.e.* $\sum_S |\widehat{p}(S)| \leqslant t$.
2) *The polynomial p approximates c in ℓ_1:* $\mathbb{E}_{x\in\{0,1\}^n}[|p(x)-c(x)|] \leqslant \varepsilon$.

Then there exists a agnostic learning algorithm for $\mathcal{C}$ that runs in time $\mathsf{poly}(t,\frac{1}{\varepsilon},n)$.

Combining this result with Theorem I.2, we get the following corollary.

Corollary I.7. *For every $\varepsilon > 0$ and $K \in \mathbb{N}$, the concept class $\mathcal{C}$ of functions whose total influence is at most K is agnostically learnable in time* $\mathsf{poly}\left(2^{O_\varepsilon(K \log K)}, \frac{1}{\varepsilon}, n\right)$

Indeed, given $\varepsilon > 0$ we apply Theorem I.2 with $\eta = \varepsilon^2$, and get the Fourier mass of f outside

$$\mathcal{S} = \left\{ S \mid \left|\widehat{f}(S)\right| \geqslant 2^{-CK \log K} \right\}$$

is at most ε^2, for some C depending on ε. Therefore the polynomial $p(x) = \sum_{S \in \mathcal{S}} \widehat{f}(S)\chi_S(x)$ is close to f in ℓ_2, i.e. $\|f - p\|_2 \leqslant \varepsilon$, and in particular $\|f - p\|_1 \leqslant \varepsilon$. Secondly, since the sum of squares of Fourier coefficients of f is at most 1, we have that $|\mathcal{S}| \leqslant 2^{2CK \log K}$, and in particular p is t-sparse for $t = 2^{2CK \log K}$.

Corollary I.7 implies that as far as polynomial time algorithms are concerned, one can agnostically learn the concept class of functions whose total influence is $O\left(\frac{\log n}{\log \log n}\right)$. This just falls short of capturing the class of polynomial size DNF formulas, which is known to be contained in the class of functions with total influence $O(\log n)$. We remark that while learning algorithms for the class of polynomial size DNF's are known [17], no agnostic learner is known. Proving the Fourier Entropy Conjecture is a known avenue towards achieving this goal [13], and we believe this avenue to be promising in light of our results. More concretely, to achieve this goal it would be enough to "shave off" the logarithmic factor from Theorem I.2 (which would also establish a conjecture of Mansour [29]).

2) Sharp thresholds: Theorem I.1 can be used to prove a quantitatively improved, nearly tight Bourgain-Kalai-like Theorem. For example, it implies that graph properties with constant variance have total influence at least $\Omega\left(\frac{\log^2 n}{(\log \log n)^2}\right)$.

Organization: In Section II we give standard tools and notions from discrete Fourier analysis. In Section III we present two important ideas that are used in the proof of our main results. In Section IV we state and prove a basic version of our main technical result, Theorem IV.1 and Corollary IV.2. In Section V we state a quantitative improvement to Corollary IV.2, which is the main technical result used in the proofs of Theorems I.1, I.2 and I.3. Due to space constraints, the proof of Corollary IV.2 as well as the derivation of Theorems I.1, I.2 and I.3, is deferred to the full version of this paper [22].

II. Preliminaries

In this section we describe the basics of Fourier analysis over the hypercube that will be needed in this paper (see [30] for a more systematic treatment).

Notations: We denote $[n] = \{1, \ldots, n\}$. Throughout the paper, $\log x$ is the natural logarithm of x. For a set $S \subseteq [n]$, we denote by $|S|$ the cardinality of S, and for $z \in \{0,1\}^n$, we denote by $|z|$ the Hamming weight of z (i.e., the number of coordinates equal to 1).

A. Fourier analysis on the hypercube

Consider the hypercube $\{0,1\}^n$ along with the uniform measure μ, and consider real-valued functions $f\colon \{0,1\}^n \to \mathbb{R}$ equipped with the inner product $\langle f, g \rangle = \mathbb{E}_{x \sim \mu}[f(x)g(x)]$. The set $\{\chi_S\}_{S \subseteq [n]}$, where $\chi_S = (-1)^{\sum_{i \in S} x_i}$ is the well-known Fourier basis that forms an orthonormal basis with respect to our inner product, thus one can expand any $f\colon \{0,1\}^n \to \mathbb{R}$ as

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x), \qquad \text{where } \widehat{f}(S) = \langle f, \chi_S \rangle.$$

The degree of a function f is $\mathsf{deg}(f) = \max_{S : \widehat{f}(S) \neq 0} |S|$. Since $\{\chi_S\}$ is an orthogonal system, we have the Parseval/Plancherel equality.

Fact II.1. *For any $f, g\colon \{0,1\}^n \to \mathbb{R}$ we have that $\langle f, g \rangle = \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{g}(S)$.*

B. Restrictions, derivatives and influences

Given a function $f\colon \{0,1\}^n \to \mathbb{R}$, a set of variables $S \subseteq [n]$ and $z \in \{0,1\}^S$, the restricted function $f_{S \to z}$ is the function from $\{0,1\}^{[n] \setminus S}$ to $\mathbb{R}$ resulting from fixing S's coordinates in x to be z. If S is a singleton $\{i\}$, we will denote this restriction by $f_{i \to z}$.

Definition II.2. *The discrete derivative of $f\colon \{0,1\}^n \to \mathbb{R}$ in direction i is a function $\partial_i f\colon \{0,1\}^{n \setminus [i]} \to \mathbb{R}$ defined by $\partial_i f(x) = \frac{1}{2}(f_{i \to 0}(x) - f_{i \to 1}(x))$.*

More generally, for a set of variables $T \subseteq [n]$, the derivative of f with respect to T is $\partial_T f\colon \{0,1\}^{[n] \setminus T} \to \mathbb{R}$ is defined by iteratively applying the derivative operator on f for each $i \in T$. Alternatively,

$$\partial_T f(x) = 2^{-|T|} \sum_{z \in \{0,1\}^T} (-1)^{|z|} f_{T \to z}(x).$$

The Fourier expansion of $\partial_T f(x)$ is $\sum_{S \supseteq T} \widehat{f}(S)\chi_{S \setminus T}(x)$.

The following definition generalizes the notion of influences to real-valued functions. We remark that for Boolean functions, it differs by a factor of 4 from the definition given in the introduction (this is done only for convenience).

Definition II.3. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ be a function. The influence of a variable $i \in [n]$ is given by $I_i[f] = \|\partial_i f\|_2^2$, and the total influence of f is defined to be $I[f] = \sum_{i=1}^{n} I_i[f]$.*

The generalized influence of a set $S \subseteq [n]$ on $f\colon \{0,1\}^n \to \mathbb{R}$ is $I_S[f] = \|\partial_S f\|_2^2$.

Using the Fourier expression for $\partial_T f$ and Parseval, we see that $I_T[f] = \sum_{S \supseteq T} \widehat{f}(S)^2$. In particular, using this formula for T's that are singletons and summing, one gets that the total influences of f can be written as $I[f] = \sum_S |S| \widehat{f}(S)^2$.

Low-degree part and low-degree influences: For $f\colon \{0,1\}^n \to \mathbb{R}$ and $d \leqslant n$, we define the degree at most d part of f, $f^{\leqslant d}$, to be

$$f^{\leqslant d}(x) = \sum_{|S| \leqslant d} \widehat{f}(S)\chi_S(x).$$

Using the Fourier formula for the total influence and Parseval, one sees that $I[f^{\leqslant d}] \leqslant d\|f\|_2^2$, and hence there are at most $\frac{d}{\tau}$ variables i that have $I_i[f^{\leqslant d}] \geqslant \tau\|f\|_2^2$. A similar property holds for generalized low-degree influences of f.

Fact II.4. *For any $v \leqslant d$ and $f\colon \{0,1\}^n \to \mathbb{R}$, one has that $\sum_{|T| \leqslant v} I_T[f^{\leqslant d}] \leqslant 2d^v \|f\|_2^2$.*

Proof: Using the Fourier formula for $I_S[f^{\leqslant d}]$, the left hand side is equal to

$$\sum_{|T| \leqslant v} I_T[f^{\leqslant d}] = \sum_{|T| \leqslant v} \sum_{\substack{|S| \leqslant d \\ S \supseteq T}} \widehat{f}(S)^2 = \sum_{|S| \leqslant d} \widehat{f}(S)^2 \sum_{\substack{T \subseteq S \\ |T| \leqslant v}} 1.$$

Fix S. The inner summation is equal to the number of subsets of S of size at most v, hence it is equal to $\binom{|S|}{0} + \ldots + \binom{|S|}{\min(v,|S|)} \leqslant \sum_{k=0}^{v} \frac{|S|^k}{k!}$. If $v \leqslant 1$, then this sum is at most $2d^v$, and if $v > 1$, we may upper bound the sum by $1 + d + \sum_{k=2}^{v} \frac{|S|^k}{k!} \leqslant 1 + d + d^v \sum_{k=2}^{\infty} \frac{1}{k!} = 1 + d + d^v(e-2) \leqslant 2d^v$. ■

C. Random restrictions

Let $f\colon \{0,1\}^n \to \mathbb{R}$ be a function, and $I \subseteq [n]$. A random restriction of f on I is the function $f_{I \to z}$ where we sample z uniformly from $\{0,1\}^I$. In our applications we will usually have two functions, $f, g\colon \{0,1\}^n \to \mathbb{R}$ and we will consider the effect of the same random restriction of them. For example, it is easy to show that for any $I \subseteq [n]$, the expected inner product of $\langle f_{I \to z}, g_{I \to z} \rangle$ over $z \in_R \{0,1\}^I$ is equal to $\langle f, g \rangle$. Another quantity associated with f, g that we will consider is the cross-total-influence.

Definition II.5. *For any $f, g\colon \{0,1\}^n \to \mathbb{R}$ and $i \in [n]$, we define the cross-influence along direction i to be $I_i[f,g] = \sqrt{I_i[f]I_i[g]}$. The cross-total-influence of f, g is given by $I[f,g] = \sum_{i=1}^{n} I_i[f,g]$.*

By Cauchy-Schwarz, one always has that $I[f,g] \leqslant \sqrt{I[f]I[g]}$. The quantity $I[f,g]$ though will be easier for us to work with inductively, and the following property will be useful for us.

Lemma II.6. *For any $f, g\colon \{0,1\}^n \to \mathbb{R}$ and $I \subseteq [n]$, we have that*

$$\mathop{\mathbb{E}}_{z \in \{0,1\}^I} [I[f_{I \to z}, g_{I \to z}]] \leqslant \sum_{i \notin I} I_i[f,g].$$

Proof: By definition, the left hand side is equal to

$$\begin{aligned}
&\mathop{\mathbb{E}}_{z \in \{0,1\}^I} \left[\sum_{i \in \overline{I}} \sqrt{I_i[f_{I \to z}]} \cdot \sqrt{I_i[g_{I \to z}]} \right] \\
&= \sum_{i \in \overline{I}} \mathop{\mathbb{E}}_{z \in \{0,1\}^I} \left[\sqrt{I_i[f_{I \to z}]} \cdot \sqrt{I_i[g_{I \to z}]} \right] \\
&\leqslant \sum_{i \in \overline{I}} \sqrt{\mathop{\mathbb{E}}_{z \in \{0,1\}^I} [I_i[f_{I \to z}]]} \cdot \sqrt{\mathop{\mathbb{E}}_{z \in \{0,1\}^I} [I_i[g_{I \to z}]]} \\
&= \sum_{i \in \overline{I}} \sqrt{I_i[f]}\sqrt{I_i[g]},
\end{aligned}$$

where the second transition is by Cauchy-Schwarz. ■

We will also need the following fact about Fourier coefficients of random restrictions.

Lemma II.7. *For any $g\colon \{0,1\}^n \to \mathbb{R}$, $I \subseteq [n]$ and $S \subseteq I$ we have that*

$$\mathop{\mathbb{E}}_{z \in \{0,1\}^{\overline{I}}} \left[\widehat{g_{\overline{I} \to z}}(S)^2 \right] = \sum_{T : T \cap I = S} \widehat{g}(T)^2.$$

Proof: The Fourier coefficient of S in $g_{\overline{I} \to z}$ is $\sum_{T \subseteq \overline{I}} \widehat{g}(S \cup T)\chi_T(z)$. Thinking of the latter as a function of z and using Parseval's equality, we get that the expectation of its square is equal to $\sum_{T \subseteq \overline{I}} \widehat{g}(S \cup T)^2$. ■

D. Hypercontractivity

We will need the hypercontractive inequality [2], [4], [15], which states that for $q \geqslant 2$, the q-norm and 2-norm of degree-d functions is comparable up to exponential factor in d.

Theorem II.8. *If $f\colon \{0,1\}^n \to \mathbb{R}$ is a function of degree at most d, and $q \geqslant 2$, then $\|f\|_q \leqslant (q-1)^{d/2}\|f\|_2$.*

III. Restrictions, partitions and degree reductions

In this section we state several lemmas that will be helpful in the proof of Theorem I.1, and we start by describing the basic motivation.

Using hypercontractivity (Theorem II.8) incurs a loss of exponential factor in the degree of the function it is applied on. This exponential-factor loss is in fact the bottleneck in KKL and Friedgut's Theorems. It is ultimately the reason it is hard to bypass the logarithmic barrier. Therefore one may search for methods by which the degree of a function could be reduced prior to applying hypercontractivity.

One natural idea is to consider random restrictions: if we pick $I \subseteq [n]$ randomly by including each element with probability $\frac{1}{2}$ in it and restrict variables outside I, then the degree of f shrinks by a factor of 2 – at least if we are willing to discard characters of small total mass from it. Another point that is often useful, is that this allows one to view the given function f as $f(y,z)$, where $y \in \{0,1\}^I$, $z \in \{0,1\}^{\bar{I}}$ and the degree of f on each one of y,z (separately) is at most $d/2$.

Sometimes, it is necessary to get a degree reduction by more than a constant factor. One can certainly decrease the size of the set of live variables I, however this introduces asymmetry between I and $\bar{I}$, and thus we may not enjoy any reduction on $\bar{I}$. The idea of *random partition* remedies this situation. To get a degree reduction by factor m, we may consider a random partition of $[n]$ into m disjoint sets, i.e. $[n] = I_1 \cup \ldots \cup I_m$ generated by including every $i \in [n]$ in each one of them with equal probability. We show that under mild conditions on d and m, given a function f there is a partition of $[n]$ into m parts and a function f' close to f in ℓ_2, such that the restrictions of f' to each one of the parts, $(f')_{\bar{I_j}\to z}$, is of degree (roughly) at most d/m.

A. Random partitions

Recall that a function $f\colon \{0,1\}^n \to \mathbb{R}$ is called degree-d homogenous if all Fourier characters in its support have size exactly d, and we would like to generalize it to functions that are "almost" degree d-homogenous. We say that $S \subseteq [n]$ is *of size d within factor α* if $\alpha d \leqslant |S| \leqslant d$, and often write it succinctly by $|S| \sim d$ (α will be clear from the context).

Definition III.1. *A function $f\colon \{0,1\}^n \to \mathbb{R}$ is called (α,d)-almost-homogenous if all Fourier characters χ_S in its support satisfy $\alpha d \leqslant |S| \leqslant d$.*

Definition III.2. *A partition of $[n]$ into m parts is $\mathcal{I} = (I_1,\ldots,I_m)$, where $I_1,\ldots,I_m$ are pairwise disjoint sets that cover $[n]$.*

As discussed earlier, a random partition into m parts is constructed by starting out with $I_1 = \ldots = I_m = \emptyset$, and then for each $i \in [n]$ choosing the part I_j to which we add i uniformly among the m-parts.

The following claim asserts that if S is of size roughly d, and we choose a random partition $\mathcal{I}$, then with high probability $|S\cap I_j|$ is roughly of size d/m for all $j \in [m]$ (for technical reasons, since the definition of "almost-homogenous" allows for α to enter only in the lower bound on S, we allow for a slack of $(1+\varepsilon)$ factor in the size too).

Lemma III.3. *Let $m,d \in \mathbb{N}$ and $\varepsilon,\alpha \in (0,1)$. If $S \subseteq [n]$ is of size d within factor α, then*

$$\Pr_{\mathcal{I}=(I_1,\ldots,I_m)}\left[\begin{matrix} S\cap I_j \textit{is of size } (1+\varepsilon)d/m \\ \textit{within factor } (1-2\varepsilon)\alpha \end{matrix}\right] \geqslant 1-2me^{-\frac{\varepsilon^2|S|}{3m}}.$$

Proof: Let $\mathcal{I} = (I_1,\ldots,I_m)$ be a random partition of $[n]$ into m parts, and for each $i \in [n]$ and $j \in [m]$ denote by $\mathbb{1}_{i\in I_j}$ the indicator function of the event that i is in I_j. Thus, for each $j \in [m]$ we may write the random variable $|S\cap I_j|$ as a sum of independent random variables $\sum\limits_{i\in S}\mathbb{1}_{i\in I_j}$. Note that by linearity of expectation, we have that its expectation is $|S|/m$, and we use Chernoff's bound to argue it is close to its expectation with high probability.

More precisely, using Chernoff's inequality for indicator random variables we have that

$$\Pr_{\mathcal{I}}\left[\left||S\cap I_j| - \frac{|S|}{m}\right| \geqslant \varepsilon\frac{|S|}{m}\right] \leqslant 2e^{-\frac{\varepsilon^2|S|}{3m}},$$

and therefore by the Union Bound the probability $\left||S\cap I_j| - \frac{|S|}{m}\right| \geqslant \varepsilon\frac{|S|}{m}$ for some $j \in [m]$ is at most m times that. ■

Next, we use the above lemma to prove that if f is almost d-homogenous, and we take a random partition $\mathcal{I}$, then on each one of the parts I_j, f is almost d/m-homogenous, provided we are willing to discard characters of small total mass from the Fourier transform of f.

More formally, let $\alpha,\varepsilon \in (0,1)$ and $d \in \mathbb{N}$ be parameters, and let $\mathcal{I} = (I_1,\ldots,I_m)$ be a partition of $[n]$. We denote by $G(\mathcal{I})$ the set of all $S \subseteq [n]$ such that $|S| \sim d$, and $S\cap I_j$ is of size $(1+\varepsilon)d/m$ within factor $(1-2\varepsilon)\alpha$ for all $j \in [m]$. In this language, the previous lemma states that provided that d is large enough in comparison to m, for each S of size d within factor α we have that $\Pr_{\mathcal{I}}[S \in G(\mathcal{I})]$ is close to 1.

Corollary III.4. *Let $m,d \in \mathbb{N}$, $\alpha,\varepsilon \in (0,1)$, and let $f,g\colon \{0,1\}^n \to \mathbb{R}$ be functions such that g is (α,d)-almost-homogenous. Then there is a partition $\mathcal{I} = (I_1,\ldots,I_m)$ such that*

$$\sum_{S\in G(\mathcal{I})}\left|\widehat{f}(S)\widehat{g}(S)\right| \geqslant \left(1-2me^{-\frac{\varepsilon^2\alpha d}{3m}}\right)\sum_{S\subseteq[n]}\left|\widehat{f}(S)\widehat{g}(S)\right|.$$

Proof: We choose a partition $\mathcal{I}$ randomly, and lower bound the expectation of the left hand side. Let $\mathbb{1}_{S\in G(\mathcal{I})}$ be the indicator random variable of S being in $G(\mathcal{I})$. By Lemma III.3, we have $\mathbb{E}\left[\mathbb{1}_{S\in\mathcal{G}(\mathcal{I})}\right] \geqslant 1-2me^{-\frac{\varepsilon^2\alpha d}{3m}}$ for each S in the support of $\widehat{g}$, and therefore

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{I}}\left[\sum_{S\in G(\mathcal{I})}\left|\widehat{f}(S)\widehat{g}(S)\right|\right] &= \sum_{S\subseteq[n]}\left|\widehat{f}(S)\widehat{g}(S)\right|\mathop{\mathbb{E}}_{\mathcal{I}}\left[\mathbb{1}_{S\in G(\mathcal{I})}\right] \\ &\geqslant \left(1-2me^{-\frac{\varepsilon^2\alpha d}{2m}}\right)\sum_{S\subseteq[n]}\left|\widehat{f}(S)\widehat{g}(S)\right|.\end{aligned}$$

In particular, there exists a choice of $\mathcal{I}$ for which the expression under the expectation on the left hand side is at least the right hand side. ■

B. Exchanging maximums and expectations

We next discuss a tool that goes in handy with partitions. Let $f\colon \{0,1\}^n \to \mathbb{R}$, let $I \subseteq [n]$ and consider the Fourier coefficients of the restricted function, i.e. for each $S \subseteq I$ consider $h_S\colon \{0,1\}^{\bar{I}} \to \mathbb{R}$ defined by $h_S(x) = \widehat{f_{\bar{I}\to x}}(S)$.

Recall that we are using restrictions (or more generally partitions) as a way to decrease the degree of the function, but eventually we want to transfer the information we got on the restrictions back to information about f. For example, if the bound we proved involves 2-norms of the Fourier coefficients of the restrictions, i.e. of h_S's, then a corresponding bound using Fourier coefficients of f can be established by Lemma II.7. The bound however could depend on the functions h_S in a more involved way, e.g. on other ℓ_p-norms of them, in which case one can often get effective bounds using hypercontractivity.

Since we are interested in the min-entropy of a function f (e.g. for Theorem I.1), we will naturally wish to understand the maximum Fourier coefficient after restriction as a function of x, $\max_S h_S(x)$, and relate it to some parameters of f. We do that in Lemma III.5.

Generalizing the above discussion, let $h_1, \ldots, h_k \colon \{0,1\}^n \to \mathbb{R}$ be functions of low-degree, and consider the following two quantities. The first one is $\mathbb{E}_{x\sim\mu} \max_i h_i(x)^2$, and in the second quantity we interchange the order of these two operations, i.e. $\max_i \mathbb{E}_{x\sim\mu} h_i(x)^2$. Note that for every x and j we have that $\max_i h_i(x)^2 \geqslant h_j(x)^2$. Taking an expectation of this inequality over $x \sim \mu$, and then maximum over j establishes that the first quantity is always larger than the second quantity. The following lemma asserts that these two quantities are polynomially related, provided that the expected value of $\sum_{i=1}^{k} h_i(x)^2$ is constant.

Lemma III.5. *If $h_1, h_2, \ldots, h_k \colon \{0,1\}^n \to \mathbb{R}$ are all of degree at most d, then*

$$\mathop{\mathbb{E}}_{x}\left[\max_{i\in[k]} h_i(x)^2\right] \leqslant 3^d \max_{i\in[k]} \|h_i\|_2 \left(\mathop{\mathbb{E}}_{x}\left[\sum_{i=1}^{k} h_i(x)^2\right]\right)^{1/2}.$$

Proof: Note that for all x, we have that $\max_{i\in[k]} h_i(x)^2 \leqslant \left(\sum_{i=1}^{k} h_i(x)^4\right)^{1/2}$, and it will be more convenient for us to upper bound the expectation of the latter function. By Jensen's inequality, its expectation is at most

$$\left(\mathop{\mathbb{E}}_{x}\left[\sum_{i=1}^{k} h_i(x)^4\right]\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{k} \mathop{\mathbb{E}}_{x}\left[h_i(x)^4\right]\right)^{\frac{1}{2}}. \tag{2}$$

Using Theorem II.8, we may upper bound $\mathbb{E}_x\left[h_i(x)^4\right]$ by $9^d \cdot \mathbb{E}_x\left[h_i(x)^2\right]^2$, and thus

$$\begin{aligned}(2) &\leqslant 3^d \left(\sum_{i\in[k]} \mathop{\mathbb{E}}_{x}\left[h_i(x)^2\right]^2\right)^{\frac{1}{2}} \\ &\leqslant 3^d \left(\max_{i\in[k]} \mathop{\mathbb{E}}_{x}\left[h_i(x)^2\right] \sum_{i\in[k]} \mathop{\mathbb{E}}_{x}\left[h_i(x)^2\right]\right)^{\frac{1}{2}},\end{aligned}$$

and using the definition of 2-norm completes the proof. ■

IV. A BASIC VERSION OF OUR MAIN TECHNICAL RESULT

In this section, we state and prove Theorem IV.1 and Corollary IV.2, which are basic, less quantitatively efficient forms of our main technical result. We find it natural though to present them along with the slightly more natural argument, and encourage the reader to read this section before moving on to Section V.

A. Proof idea

Before we prove (or even state) our main technical result, we begin with an informal overview of the idea. We start with presenting the (one-line) proof of the KKL Theorem. Let $f \colon \{0,1\}^n \to \{0,1\}$, $d \in \mathbb{N}$, and $g = f^{\leqslant d} - \widehat{f}(\emptyset)$; we have that:

$$\begin{aligned}\langle f, g\rangle \leqslant \sum_{i=1}^{n} \langle \partial_i f, \partial_i(f^{\leqslant d})\rangle &\leqslant \sum_{i=1}^{n} \|\partial_i f\|_{4/3} \|\partial_i(f^{\leqslant d})\|_4 \\ &\leqslant \sqrt{3}^d \sum_{i=1}^{n} \|\partial_i f\|_{4/3} \|\partial_i(f^{\leqslant d})\|_2 \\ &\leqslant 2\sqrt{3}^d I[f] \max_i I_i[f^{\leqslant d}]^{1/4},\end{aligned}$$

where in the second inequality we used Hölder's inequality and in the third inequality we used Theorem II.8. In the last inequality, we upper bounded $\|\partial_i(f^{\leqslant d})\|_2$ by $\|\partial_i f\|_2^{1/2} \cdot \|\partial_i(f^{\leqslant d})\|_2^{1/2} = I_i[f]^{1/4} \|\partial_i(f^{\leqslant d})\|_2^{1/2}$, and used the Booleanity of f to bound $\|\partial_i f\|_{4/3} \leqslant 2 I_i[f]^{3/4}$. Hence, this inequality gives us very good bounds on the weight f has on its low-degrees, provided that all of its low degree influences are small. This raises two questions:

1) What bounds could be proved if we know that there are only few influential variables?
2) Can we improve on this bound if we know stronger information about f, e.g. that its generalized low-degree influences, $I_S[f^{\leqslant d}]$, are all very small?

For the first question, a standard bound proceeds by handling characters S that consist only of influential variables separately (similar to Lemma IV.3 below). In essence, the bound says that if there are at most T variables with influence at least δ, then there are at most T^d characters consisting only of these influential variables, and one gets the bound $T^d \max_S \left|\widehat{f}(S)\widehat{g}(S)\right| + \sqrt{3}^d I[f]\delta^{1/4}$, which in this case is just

$$T^d \max_S \widehat{f}(S)^2 + \sqrt{3}^d I[f]\delta^{1/4}. \tag{3}$$

The second question is more interesting, and the obvious naive attempt one may first have quickly fails.[4]

[4]Namely, the attempt is to run the proof of the KKL theorem with the generalized derivatives instead of standard derivatives. This fails since the sum of the generalized derivatives in general may be much higher than the total influence: if we consider say, order v derivatives, then a character S would be counted by $\binom{|S|}{v}$ generalized influences.

One key insight in [6], is that a better way to use the information on generalized influence is by relating them to Fourier coefficients of random restrictions (and importantly, to something slightly stronger, see Lemma II.7). For a function $f\colon \{0,1\}^n \to \{0,1\}$, a set $I \subseteq [n]$ and a subset $S \subseteq I$, the expected Fourier coefficient squared, $\widehat{f_{\bar{I}\to z}}(S)^2$, is at most $I_S[f]$; also, the degree of a random restriction (roughly speaking) of $f^{\leqslant d}$ is significantly smaller than d. Thus, by first applying a random restriction, and then using Inequality 3 (the "KKL bound"), one may expect to get a meaningful bound for the second question above — this is indeed the case. An important technical point is that one may "switch" the order of maximum and expectation in this argument, which is where Lemma III.5 comes in handy.

The above argument gives a stronger bound than Inequality (3), provided all of the generalized influence of f (of some order) are very small; this is very much analogous to the KKL Theorem we started with! Again, the following question arises: can we base on this result a similar bound to Inequality (3) in case we know f only has a few noticeable generalized influences? This is the next step in the argument, and is established in a similar manner to the way we established Inequality (3).

The final statement, Theorem IV.1 below, is the outcome of applying this idea inductively. There are several technical points omitted from the above description that need to be taken into account to make it precise. To get a strong enough statement, the set of live variables I should be chosen randomly, but at the same time the degree of $f^{\leqslant d}$ under random restrictions has to decrease. While this can probably be done, it is likely to be messy, and we bypass it by using random-partitions from Section III-A. We then discard from the Fourier expansion of $f^{\leqslant d}$ characters S that remain large with respect to one of the parts (and argue they do not contribute too much to $\langle f, f^{\leqslant d}\rangle$). Note that after discarding, it will no longer be true that we are working with a function and its low degree part, so to facilitate induction we work with the two-function version of the problem instead. We then look at each part I_j of the partition $\mathcal{I}$, consider random restrictions of the discarded version of the low-degree part of f and apply the induction hypothesis on them (note that these restrictions also decrease the degree of the function considerably, as the partition $\mathcal{I}$ is picked according to Corollary III.4). A slightly more detailed overview of the inductive step is given in Section IV-D.

B. Statement of the main technical result

In this section, we prove our main technical result. In the following statement, we have $\varepsilon > 0$, and an increasing sequence of integers $d_0, \ldots, d_k$, and we will be interested in S that are of size d_k within factor α; it will be convenient for us to write it more succinctly as $|S| \sim d_k$.

In the statement below, one should think of g (roughly) as $f^{\leqslant d}$, and the goal is to prove that if f has no large Fourier coefficients, then $\langle f, g\rangle = o(\|f\|_2^2)$, hence f is concentrated on high degrees and in particular $I[f] \geqslant (1-o(1))d\|f\|_2^2$. The more general statement with general g is crucial for the inductive proof to go through (as hinted in the overview above).

Theorem IV.1. *Let $k \in \mathbb{N}$, $\alpha, \varepsilon \in (0,1)$, let $d_0, d_1, \ldots, d_k$ be an increasing sequence such that $d_0 = 1$ and for each $3 \leqslant j \leqslant k$ we have $d_{j-1} \geqslant \frac{3}{\alpha\varepsilon^2(1-2\varepsilon)^k}\log(4(1+\varepsilon)d_j/d_{j-1})$, and let $0 < \delta_k \leqslant \ldots \leqslant \delta_1$. Then there are C_1, C_2, C_3 specified below, such that the following holds.*

If $f\colon \{0,1\}^n \to \{0,1\}$ is a Boolean function, and $g\colon \{0,1\}^n \to \mathbb{R}$ is (α, d_k)-almost-homogenous, then

$$\langle f, g\rangle \leqslant C_1 \cdot \max_{|S|\sim d_k} \left|\widehat{f}(S)\widehat{g}(S)\right| + C_2 \cdot I[f,g] + C_3 \cdot \|g\|_2\|f\|_2. \tag{4}$$

$$C_1 = 2^k d_k^{(1+\varepsilon)d_k}\left(\frac{2}{\delta_k}\right)^{\frac{(1+\varepsilon)d_k}{d_{k-1}}}, \quad C_2 = 2^k\delta_1^{\frac{1}{8}}3^{\frac{d_1}{4}},$$

$$C_3 = 2^k(1+\varepsilon)^k d_k \sum_{j=1}^{k-1} d_j^{(1+\varepsilon)d_j}\left(\frac{2}{\delta_j}\right)^{\frac{(1+\varepsilon)d_j}{d_{j-1}}}\sqrt{3}^{d_{j+1}}\delta_{j+1}^{\frac{1}{4}}.$$

The amount of parameters in the above statement, as well as the formulas for C_1, C_2, C_3, make the above statement incomprehensible; this form is very convenient for the inductive proof to go through. Once it has been established, one can make a particular choice of the parameters that is typically useful, yielding the following corollary.

Corollary IV.2. *Let $\alpha \in (0,1)$, $\varepsilon \in (0,1/2)$ and let $d \in \mathbb{N}$, $\delta > 0$ be such that $d^\varepsilon \geqslant \frac{100}{\alpha\varepsilon}\log d$ and $\delta \leqslant 2^{-4d^\varepsilon}$. If $f\colon \{0,1\}^n \to \{0,1\}$ is a Boolean function, and $g\colon \{0,1\}^n \to \mathbb{R}$ is (α, d)-almost-homogenous, then*

$$\langle f, g\rangle \leqslant \delta^{-2^{O(1/\varepsilon)}d} \cdot \max_{|S|\sim d}\left|\widehat{f}(S)\widehat{g}(S)\right| + 2^{O(1/\varepsilon)}\delta^{1/40} \cdot I[f,g] + 2^{O(1/\varepsilon)}\delta^{d^\varepsilon} \cdot \|g\|_2\|f\|_2.$$

Proof: Assume $1/\varepsilon$ is an integer (otherwise we may replace ε with some ε' such that $\varepsilon \leqslant \varepsilon' \leqslant 2\varepsilon$ for which $1/\varepsilon'$ is an integer), and set $k = \frac{1}{\varepsilon}$. Choose $d_j = d^{j/k}$ for all $j = 0, 1, \ldots, k$, and note that by the lower bound on d, we have $d_{j-1} \geqslant \frac{3}{\alpha\varepsilon^2(1-2\varepsilon)^k}\log(4(1+\varepsilon)d_j/d_{j-1})$ for all $j \geqslant 3$. Next, we choose $\delta_1 = \delta$ and $\delta_{j+1} = \delta_j^{80d^\varepsilon}$ for all $j \geqslant 1$. We apply Theorem IV.1 with these parameters to upper bound $\langle f, g\rangle$, and get that $\langle f, g\rangle \leqslant C_1 \max_{|S|\sim d_k}\left|\widehat{f}(S)\widehat{g}(S)\right| + C_2 \cdot I[f,g] + C_3 \cdot \|g\|_2\|f\|_2$ for C_1, C_2, C_3 as in the statement of Theorem IV.1, and next we give simpler upper bounds on C_1, C_2, C_3 for our specific choice of parameters.

Unraveling the definition of δ_{j+1}, we see that $\delta_{j+1} = \delta^{(80d^\varepsilon)^j}$, therefore $C_1 \leqslant d^{O(d)}\delta^{-\exp(O(1/\varepsilon))d}$, and since $\delta \leqslant 2^{-d^\varepsilon} \leqslant 2^{-\log d} = 1/d$, we get that the $d^{O(d)}$ factor can be absorbed into the second factor. For C_2, we see that $C_2 \leqslant 2^{O(1/\varepsilon)}\delta^{1/8}3^{d^\varepsilon/4} \leqslant 2^{O(1/\varepsilon)}\delta^{1/40}$ using the upper bound on δ. Finally, for C_3, consider the jth summand; note that $d_j^{d_j} \leqslant$

$2^{d_j \log d} \leqslant 2^{d_j d^{\varepsilon}} = 2^{d_{j+1}}$, and that $\delta_{j+1}^{1/4} \leqslant \delta_j^{20 d^{\varepsilon}}$. Thus,

$$\begin{aligned}
C_3 &\leqslant 2^k (1+\varepsilon)^k d \sum_{j=1}^{k-1} 12^{d_{j+1}} \delta_j^{16 d^{\varepsilon}} \\
&\leqslant 2^k (1+\varepsilon)^k d \sum_{j=1}^{k-1} 12^{d^{(j+1)\varepsilon}} \delta^{16 d^{j\varepsilon}} \\
&\leqslant 2^k (1+\varepsilon)^k d \sum_{j=1}^{k-1} \delta^{8 d^{j\varepsilon}},
\end{aligned}$$

where in the last inequality we used the fact that $\delta \leqslant 2^{-d^{\varepsilon}}$. We bound the sum by k times the maximum summand and get that $C_3 \leqslant 2^{O(1/\varepsilon)} d \delta^{8 d^{\varepsilon}} \leqslant 2^{O(1/\varepsilon)} \delta^{d^{\varepsilon}}$. ■

C. Base case

The base case $k = 1$ of Theorem IV.1 is an easy consequence of Lemma IV.3 below.

Lemma IV.3. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function and let $g\colon \{0,1\}^n \to \mathbb{R}$ be of degree at most d. Then for all $\delta > 0$ we have that*

$$\langle f, g\rangle \leqslant \left(\frac{d}{\delta}\right)^d \max\left|\widehat{f}(S)\widehat{g}(S)\right| + 2\cdot \delta^{1/8} 3^{d/4} I[f,g].$$

Proof: Note that without loss of generality, we may assume that for every S, the signs of the coefficients of S in f and g are the same: indeed, for any other S we may change the sign of $\widehat{g}(S)$, leave the right hand side unchanged and only increase the left hand side (as evident from Parseval/ Plancherel).

Denote by $\mathsf{Inf}[f]$ the set of variables i such that $I_i[f^{\leqslant d}] \geqslant \delta$. Writing $\langle f, g\rangle = \sum\limits_{S} \widehat{f}(S)\widehat{g}(S)$, we partition the sum on the right hand side to $S \subseteq \mathsf{Inf}[f]$ (i.e. only contain variables with high low-degree influence), and the rest. Clearly, we have

$$\begin{aligned}
\langle f, g\rangle &= \sum_{S} \widehat{f}(S)\widehat{g}(S) \\
&\leqslant \underbrace{\sum_{S \subseteq \mathsf{Inf}[f]} \widehat{f}(S)\widehat{g}(S)}_{(I)} + \sum_{i \notin \mathsf{Inf}[f]} \underbrace{\sum_{S \ni i} \widehat{f}(S)\widehat{g}(S)}_{(II)}.
\end{aligned}$$

Note that $|\mathsf{Inf}[f]| \leqslant \frac{d}{\delta}$, so the total number of summands in (I) is at most $\left(\frac{d}{\delta}\right)^d$, and we get that

$$(I) \leqslant \left(\frac{d}{\delta}\right)^d \max_{S} \widehat{f}(S)\widehat{g}(S).$$

We now proceed to upper bound (II) for each i separately. Fix $i \notin \mathsf{Inf}[f]$. Since the sum is only supported on S of size at most d, we may replace $\widehat{f}(S)$ in that sum with $\widehat{f^{\leqslant d}}(S)$ and not change it. Hence, we get that

$$(II) = \langle \partial_i f^{\leqslant d}, \partial_i g\rangle,$$

and using the Cauchy-Schwarz inequality (II) is upper bounded by $\|\partial_i f^{\leqslant d}\|_2 \|\partial_i g\|_2$. We wish to upper bound the first multiplicand further, and for that we note that $\|\partial_i f^{\leqslant d}\|_2^2 = \langle \partial_i f^{\leqslant d}, \partial_i f^{\leqslant d}\rangle = \langle \partial_i f^{\leqslant d}, \partial_i f\rangle$ and then use Hölder's inequality with powers $(4, 4/3)$ to get that

$$\begin{aligned}
\langle \partial_i f^{\leqslant d}, \partial_i f\rangle &\leqslant \|\partial_i f^{\leqslant d}\|_4 \|\partial_i f\|_{4/3} \\
&\leqslant \sqrt{3}^d \|\partial_i f^{\leqslant d}\|_2 \|\partial_i f\|_{4/3},
\end{aligned}$$

where in the last inequality we used Theorem II.8. Since $i \notin \mathsf{Inf}[f]$, we have that $\|\partial_i f^{\leqslant d}\|_2 \leqslant \delta^{1/4} \cdot \|\partial_i f^{\leqslant d}\|_2^{1/2}$, which by Parseval is at most $\delta^{1/4} \cdot I_i[f]^{1/4}$. To upper bound $\|\partial_i f\|_{4/3}$, as $\partial_i f$ is $\{0, 1/2, -1/2\}$-valued, we have that $\|\partial_i f\|_{4/3} \leqslant 2 I_i[f]^{3/4}$. Combining the two bounds and taking square root, we get that $\|\partial_i f^{\leqslant d}\|_2 \leqslant 2 \cdot 3^{d/4} \delta^{1/8} I_i[f]^{1/2}$, and therefore $(II) \leqslant 2 \cdot 3^{d/4} \delta^{\frac{1}{8}} \|\partial_i f\|_2 \|\partial_i g\|_2$. ■

D. Proof of inductive step

In this section, we prove Theorem IV.1 by induction on k. The base case $k = 1$ follows from Lemma IV.3, noting that since in our case g is (α, d_1)-almost-homogenous, the maximum could be restricted to $|S| \sim d_1$.

Let $k > 1$, assume the statement holds for all $j < k$ and prove it for k. To simplify notation, we recall that by $|S| \sim d_k$ we mean that $\alpha d_k \leqslant |S| \leqslant d_k$; also, for $j \leqslant k-1$, we say that $|S| \sim d_j$ if $(1-2\varepsilon)\alpha d_j \leqslant |S| \leqslant d_j$.

Proof overview of the inductive step: Let f, g be functions as in the statement of the theorem, and consider a partition $\mathcal{I} = (I_1, \ldots, I_m)$ into $m = d_k/d_{k-1}$ parts as in Lemma III.3; this partition could be thought of as random. We discard from g characters χ_S for which $|S \cap I_j| \gg d_{k-1}$ for some $j \in [m]$. Thus, thinking of g as a function of only one of the parts, say I_j, its degree is at most d_{k-1}. Our goal is to charge S's that contribute to $\langle f, g\rangle = \sum\limits_{S} \widehat{f}(S)\widehat{g}(S)$ to the various random restrictions $f_{\bar{I_j} \to z}, g_{\bar{I_j} \to z}$ for $j \in [m]$, where z is a random setting outside the coordinates of I_j, and bound the contribution to the random restrictions using the induction hypothesis.

As hinted earlier, if all of the generalized influences of g corresponding to subsets of I_j are small, then the induction hypothesis allows us to establish a good bound on $\langle f, g\rangle$ by writing it as $\mathbb{E}_z\left[\langle f_{\bar{I_j} \to z}, g_{\bar{I_j} \to z}\rangle\right]$. To see that, we focus on the most problematic term that arises from the induction hypothesis involving the maximum over Fourier coefficients (the rest are significantly easier to handle), i.e.

$$\mathop{\mathbb{E}}_{z \in \{0,1\}^{\bar{I_j}}} \left[\max_{S}\left|\widehat{f_{\bar{I_j} \to z}}(S)\widehat{g_{\bar{I_j} \to z}}(S)\right|\right].$$

To prove a good upper bound on this term, one uses Cauchy-Schwarz and then Lemma III.5 to exchange the maximum and expectation, which results in a bound depending on the generalized influences of g.

However, g could of course have large generalized influences on I_j. Thus, we take g_j to be the part of the

Fourier transform of g that consists only of characters χ_S for which $S \cap I_j$ has small generalized influences (if for some S we have that $S \cap I_j$ is non-influential for more than a single j, we choose one such j arbitrarily and include χ_S in g_j). Thus, we are able to upper bound $\langle f, g_j \rangle$ successfully using the above strategy. Subsequently we can decompose g as $\sum\limits_{j\in[m]} g_j + E$, where E consists of characters χ_S for which $S \cap I_j$ is influential for all $j \in [m]$. The task then amounts to upper bounding $\langle f, E \rangle$, and for that we use the crude upper bound $\left|\mathsf{supp}(\widehat{E})\right| \max_S \left|\widehat{f}(S)\widehat{E}(S)\right|$, which is sufficient (using Fact II.4 to upper bound the size of the support of $\widehat{E}$).

We now move on to the formal proof.

Formal proof: As before, we may assume without loss of generality that for all Fourier coefficients S, the signs of $\widehat{f}(S)$ and $\widehat{g}(S)$ are the same.

Using Corollary III.4 with $m = (1+\varepsilon)\frac{d_k}{d_{k-1}}$, $d = d_k$, α and ε we may find a partition $\mathcal{I}$ as in the corollary; let $G(\mathcal{I})$ be the set of all $S \subseteq [n]$ such that for all $j \in [m]$, $S \cap I_j$ is of size $(1+\varepsilon)d_k/m = d_{k-1}$ within factor $(1-2\varepsilon)\alpha$.

Define $\tilde{g}\colon \{0,1\}^n \to \mathbb{R}$ by $\tilde{g} = \sum_{S\in G(\mathcal{I})} \widehat{g}(S)\chi_S$. In terms of $\tilde{g}$, Corollary III.4 amounts to saying that

$$\langle f, \tilde{g} \rangle \geqslant \left(1 - 2me^{-\frac{\varepsilon^2 \alpha d}{3m}}\right) \langle f, g \rangle,$$

and by the condition relating d_k and d_{k-1} we get that the factor on the right hand side is at least $\frac{1}{2}$, and therefore $\langle f, g \rangle \leqslant 2\langle f, \tilde{g} \rangle$ so it is enough to upper bound the inner product of f and $\tilde{g}$.

Let $\mathcal{T}$ be the set of $T \subseteq [n]$ of size d_{k-1} within factor $(1-\varepsilon)\alpha$ that have large generalized influence in $\tilde{g}$, i.e. such that $I_T[\tilde{g}] \geqslant \delta_k \|\tilde{g}\|_2^2$. Since $\tilde{g}$ has degree at most d_k, by Fact II.4 we get that

$$\sum_{|T|\sim d_{k-1}} I_T[\tilde{g}] \leqslant \sum_{|T|\leqslant d_{k-1}} I_T[\tilde{g}] \leqslant 2d_k^{d_{k-1}} \|g\|_2^2,$$

hence $|\mathcal{T}| \leqslant \frac{2d_k^{d_{k-1}}}{\delta_k}$. Writing $\langle f, \tilde{g} \rangle = \sum\limits_S \widehat{f}(S)\widehat{\tilde{g}}(S)$, we partition the sum on the right hand side into two parts: (I) those S that satisfy that $S \cap I_j$ is in $\mathcal{T}$ for all $j \in [m]$, and (II) those S such that $S \cap I_j$ is not in $\mathcal{T}$ for some $j \in [m]$. Denote by $\mathcal{S}_j$ the set of S such that $S \cap I_j \notin \mathcal{T}$, and by $\mathcal{B}$ the set of S that are outside $\mathcal{S}_1 \cup \ldots \cup \mathcal{S}_m$. Then we have

$$\langle f, \tilde{g} \rangle \leqslant \underbrace{\sum_{S\in\mathcal{B}} \widehat{f}(S)\widehat{\tilde{g}}(S)}_{(I)} + \underbrace{\sum_{j=1}^{m} \sum_{S\in\mathcal{S}_j} \widehat{f}(S)\widehat{\tilde{g}}(S)}_{(II)}, \tag{5}$$

and we upper bound each sum separately.

Upper bounding (I)*:* Clearly, (I) is at most $|\mathcal{B}| \cdot \max_{|S|\sim d_k} \widehat{f}(S)\widehat{\tilde{g}}(S)$ (as the sum is only supported on $|S| \sim d_k$ by the condition on g). To bound the size of $\mathcal{B}$, note that the map $S \to (S \cap I_1, \ldots, S \cap I_m)$ is a bijection from $\mathcal{B}$ to $\mathcal{T}^m$, hence we have that

$$|\mathcal{B}| \leqslant |\mathcal{T}|^m \leqslant \left(\frac{2d_k^{d_{k-1}}}{\delta_k}\right)^m = d_k^{(1+\varepsilon)d_k} \left(\frac{2}{\delta_k}\right)^{\frac{(1+\varepsilon)d_k}{d_{k-1}}},$$

which is at most $\frac{1}{2}C_1(k)$. Therefore, the contribution from (I) is upper bounded by the first term on the right hand side in (4). Thus, to complete the proof, it is enough to upper bound the contribution from (II) by the other two terms in the right hand side of (4), and to do so we use the induction hypothesis.

Upper bounding (II)*:* We upper bound the sum corresponding to each $j \in [m]$ separately. Fix j, write $J = \cup_{j'\neq j} I_{j'}$, and $\tilde{g}_j = \sum\limits_{S\in\mathcal{S}_j} \widehat{g}(S)\chi_S$. Note that

$$\mathbb{E}_z\left[\langle f_{J\to z}, (\tilde{g}_j)_{J\to z}\rangle\right] = \langle f, \tilde{g}_j \rangle = \sum_{S\in\mathcal{S}_j} \widehat{f}(S)\widehat{g}(S),$$

and therefore to upper bound (II), it is enough to upper bound the inner product of random restrictions of f and $\tilde{g}_j$ on J. We note that the important point here is that these restrictions lower the degree of $\tilde{g}_j$ from $\sim d_k$ to $\sim d_{k-1}$ (since it is only supported on S such that $|S \cap I_j| \sim d_{k-1}$), and hence we expect to get useful information from the inductive hypothesis on these restrictions.

More precisely, note that for every $z \in \{0,1\}^J$ the function $f_{J\to z}$ is Boolean and the function $(\tilde{g}_j)_{J\to z}$ is $((1-2\varepsilon)\alpha, d_{k-1})$-approximately homogenous. Therefore, we may apply the induction hypothesis with parameters $k-1$, $(1-2\varepsilon)\alpha$, ε, $d_0, \ldots, d_{k-1}$ and $\delta_{k-1}, \ldots, \delta_1$ on these functions, to get that

$$\begin{aligned}
&\langle f_{J\to z}, (\tilde{g}_j)_{J\to z}\rangle \\
&\leqslant C_1(k-1) \underbrace{\max_{|S|\sim d_{k-1}} \left|\widehat{f_{J\to z}}(S)\right| \left|\widehat{(\tilde{g}_j)_{J\to z}}(S)\right|}_{(III)} \\
&\quad + C_2(k-1)\, \underbrace{I[f_{J\to z}, (\tilde{g}_j)_{J\to z}]}_{(IV)} \\
&\quad + C_3(k-1)\, \underbrace{\|(\tilde{g}_j)_{J\to z}\|_2 \|f_{J\to z}\|_2}_{(V)}.
\end{aligned} \tag{6}$$

Here again, we denote

$$C_1(r) = 2^r \cdot d_r^{(1+\varepsilon)d_r} \left(\frac{2}{\delta_r}\right)^{\frac{(1+\varepsilon)d_r}{d_{r-1}}}, C_2(r) = 2^r \cdot \delta_1^{\frac{1}{8}} 3^{\frac{d_1}{4}},$$

$$C_3(r) = 2^r (1+\varepsilon)^r d_r \sum_{\ell=1}^{r-1} \left(\frac{2d_\ell^{d_{\ell-1}}}{\delta_\ell}\right)^{\frac{(1+\varepsilon)d_\ell}{d_{\ell-1}}} \sqrt{3}^{d_{\ell+1}} \delta_{\ell+1}^{\frac{1}{4}},$$

and we wish to upper bound the expectation of the left hand side. We bound each one of them separately.

Upper bounding the contribution of (V): By Cauchy-Schwarz we have that

$$\mathop{\mathbb{E}}_{z}[(V)] = \mathop{\mathbb{E}}_{z}[\|(\tilde{g}_j)_{J\to z}\|_2\|f_{J\to z}\|_2]$$
$$\leqslant \sqrt{\mathop{\mathbb{E}}_{z}[\|(\tilde{g}_j)_{J\to z}\|_2^2]\mathop{\mathbb{E}}_{z}[\|f_{J\to z}\|_2^2]} = \|\tilde{g}_j\|_2\|f\|_2 \leqslant \|g\|_2\|f\|_2.$$

Upper bounding the contribution of (IV): Using Lemma II.6, we have that

$$\mathop{\mathbb{E}}_{z}[(IV)] = \mathop{\mathbb{E}}_{z}[I[f_{J\to z},(\tilde{g}_j)_{J\to z}]] \leqslant \sum_{i\notin J} I_i[f,\tilde{g}_j],$$

which is at most $\sum\limits_{i\in I_j} I_i[f,g]$ since $I_i[\tilde{g}_j] \leqslant I_i[g]$.

Upper bounding the contribution of (III): To upper bound the expectation of (III), we first use Cauchy-Schwarz inequality to see that

$$\mathop{\mathbb{E}}_{z}[(III)]$$
$$\leqslant \sqrt{\mathop{\mathbb{E}}_{z}\left[\max_{|S|\sim d_{k-1}} \widehat{f_{J\to z}}(S)^2\right]}\sqrt{\mathop{\mathbb{E}}_{z}\left[\max_{|S|\sim d_{k-1}} \widehat{(\tilde{g}_j)_{J\to z}}(S)^2\right]}.$$

For each z and each $|S|\sim d_{k-1}$, by Parseval we have that $\widehat{f_{J\to z}}(S)^2 \leqslant \|f_{J\to z}\|_2^2$, hence, the first multiplicand is bounded by $\|f\|_2$.

For the second multiplicand we appeal to Lemma III.5: let us think of the indices therein as being subsets S, and define $h_S(z) = \widehat{(\tilde{g}_j)_{J\to z}}(S)$; note that since g is of degree at most d_k, we get that each h_S also has degree at most d_k. Thus, applying Lemma III.5 we get that

$$\mathop{\mathbb{E}}_{z}\left[\max_{S} h_S(z)^2\right] \leqslant 3^{d_k}\max_{S}\|h_S\|_2\left(\mathop{\mathbb{E}}_{z}\left[\sum_{S} h_S(z)^2\right]\right)^{1/2}.$$

Note that by Parseval,

$$\mathop{\mathbb{E}}_{z}\left[\sum_{S} h_S(z)^2\right] = \mathop{\mathbb{E}}_{z}\left[\|(\tilde{g}_j)_{J\to z}\|_2^2\right] = \|\tilde{g}_j\|_2^2 \leqslant \|g\|_2^2,$$

and therefore we conclude that $\mathbb{E}_z[(III)] \leqslant \sqrt{3}^{d_k}\|g\|_2^{1/2}\|f\|_2\max_S\|h_S\|_2^{1/2}$. Fix S that attains this maximum; using Lemma II.7, we see that $\|h_S\|_2^2 = \sum\limits_{T:T\cap I_j=S}\widehat{\tilde{g}_j}(T)^2$. Note that if $S\in\mathcal{T}$, then the above sum would be empty (since we do not include a character T in $\tilde{g}_j$ if $T\cap I_j$ has large generalized influence in g), and the sum would be 0. Hence we may assume that $S\notin\mathcal{T}$, and therefore this sum is at most $I_S[g]\leqslant\delta_k\|g\|_2^2$. Combining, we get that $\mathbb{E}_z[(III)] \leqslant \sqrt{3}^{d_k}\delta_k^{1/4}\|g\|_2\|f\|_2$.

Combining the bounds for $(III),(IV),(V)$: Plugging the bounds into (6) and summing over $j\in[m]$, we get that

$$(II) = \sum_{j=1}^{m}\mathop{\mathbb{E}}_{z}[\langle f_{J\to z},(\tilde{g}_j)_{J\to z}\rangle]$$
$$\leqslant m\cdot C_1(k-1)\sqrt{3}^{d_k}\delta_k^{1/4}\|g\|_2 + C_2(k-1)I[f,g]$$
$$+ m\cdot C_3(k-1)\|g\|_2\|f\|_2.$$

Consider the first and the third terms on the right hand side. Note that

$$m\cdot C_1(k-1)\sqrt{3}^{d_k}\delta_k^{1/4} + m\cdot C_3(k-1) \leqslant \frac{1}{2}C_3(k),$$

as well as that $C_2(k-1) = \frac{1}{2}C_2(k)$. Therefore the above inequality implies that $(II) \leqslant \frac{1}{2}C_2(k)I[f,g] + \frac{1}{2}C_3(k)\|g\|_2\|f\|_2$. Plugging this, as well as the bound we have on (I), into (5), we get that

$$\langle f,\tilde{g}\rangle \leqslant \frac{1}{2}C_1(k)\max_{|S|\sim d_k}\widehat{f}(S)\widehat{\tilde{g}}(S)$$
$$+\frac{1}{2}C_2(k)I[f,g] + \frac{1}{2}C_3(k)\|g\|_2\|f\|_2,$$

and since $\langle f,g\rangle \leqslant 2\langle f,\tilde{g}\rangle$, the proof is complete. ∎

V. Improving the result

In the full version of this paper, we prove the following quantitative improvement of Corollary IV.2.

Corollary V.1. *Let $\alpha\in(0,1]$, $d\in\mathbb{N}$ and $\delta>0$ be such that $\delta\leqslant\frac{1}{d^{96000/\sqrt{\alpha}}}$. If $f\colon\{0,1\}^n\to\{0,1\}$, and $g\colon\{0,1\}^n\to\mathbb{R}$ is (α,d)-almost-homogenous, then*

$$\langle f,g\rangle \leqslant \delta^{-10^5 d}\cdot\|f\|_2^{\frac{3}{4}}\|g\|_2^{\frac{3}{4}}\cdot\max_{|S|\sim d}\left|\widehat{f}(S)\widehat{g}(S)\right|^{\frac{1}{4}}$$
$$+\delta^{1/16}\cdot I[f,g] + \delta\cdot\|f\|_2\|g\|_2.$$

VI. Conclusion

As remarked in the introduction, a variant of our main result also exists for bounded functions. The proof is almost identical, except that one redefines the cross total influence as $I[f,g] = \sum\limits_{i=1}^{n}\|\partial_i f\|_2^{1/4}\|\partial_i f\|_{4/3}^{1/2}\|\partial_i g\|_2$. It would be interesting to combine our techniques with other techniques, such as variants of Håstad's switching lemma [16], to establish the Fourier entropy conjecture for AC_0-functions. It would also be interesting to find an analog of the technique presented herein in the p-biased cube for $p=o(1)$.

Acknowledgement

We thank Gil Kalai for valuable conversations over the years and helpful comments. We also thank Yuval Filmus, Nathan Keller and Subhash Khot for comments about an earlier version of this manuscript.

This work was done while author Minzer was a member in the Institute for Advanced Study, Princeton, supported partially by NSF grant DMS-1638352 and Rothschild Fellowship. Author Safra was supported by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (Grant agreement No. 835152).

References

[1] Srinivasan Arunachalam, Sourav Chakraborty, Michal Koucký, Nitin Saurabh, and Ronald de Wolf. Improved bounds on Fourier entropy and min-entropy. *arXiv preprint arXiv:1809.09819*, 2018.

[2] William Beckner. Inequalities in Fourier Analysis. *Annals of Mathematics*, 102(1):159–182, 1975.

[3] M. Ben-Or and N. Linial. Collective coin flipping, robust voting schemes and minima of banzhaf values. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 408–416, Oct 1985.

[4] Aline Bonami. Étude des coefficients de Fourier des fonctions de $l^p(g)$. *Annales de l'Institut Fourier*, 20(2):335–402, 1970.

[5] J. Bourgain. On the distribution of the Fourier spectrum of Boolean functions. *Israel J. of Math.*, (131):269–276, 2002.

[6] Jean Bourgain and Gil Kalai. Influences of variables and threshold intervals under group symmetries. *Geometric and Functional Analysis*, 7(3):438–461, 1997.

[7] Sourav Chakraborty, Raghav Kulkarni, Satyanarayana V Lokam, and Nitin Saurabh. Upper bounds on Fourier entropy. *Theoretical Computer Science*, 654:92–112, 2016.

[8] Irit Dinur and Ehud Friedgut. Intersecting families are essentially contained in juntas. *Combinatorics, Probability & Computing*, 18(1-2):107–122, 2009.

[9] Irit Dinur and Samuel Safra. On the Hardness of Approximating Minimum Vertex Cover. *Annals of Mathematics*, 162(1):439–485, 2005.

[10] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[11] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American mathematical Society*, 124(10):2993–3002, 1996.

[12] Oded Goldreich and Leonid A. Levin. A hard-core predicate for all one-way functions. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washigton, USA*, pages 25–32, 1989.

[13] Parikshit Gopalan, Adam Kalai, and Adam R Klivans. A query algorithm for agnostically learning dnf?.

[14] Parikshit Gopalan, Adam Tauman Kalai, and Adam R. Klivans. Agnostically learning decision trees. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 527–536, 2008.

[15] Leonard Gross. Logarithmic Sobolev Inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[16] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing, May 28-30, 1986, Berkeley, California, USA*, pages 6–20, 1986.

[17] Jeffrey C. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *J. Comput. Syst. Sci.*, 55(3):414–440, 1997.

[18] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *FOCS 1988*, pages 68–80, 1988.

[19] Gil Kalai. The entropy/influence conjecture. https://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/, 2007. [Online; accessed 26-October-2019].

[20] Michael J Kearns, Robert E Schapire, and Linda M Sellie. Toward efficient agnostic learning. *Machine Learning*, 17(2-3):115–141, 1994.

[21] Nathan Keller and Noam Lifshitz. The junta method for hypergraphs and the Erdős-Chvátal simplex conjecture, 2017.

[22] Esty Kelman, Guy Kindler, Noam Lifshitz, Dor Minzer, and Muli Safra. Towards a proof of the fourier–entropy conjecture? *CoRR*, abs/1911.10579, 2019.

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

[24] Subhash Khot and Assaf Naor. Nonembeddability theorems via Fourier analysis. In *FOCS 2005*, pages 101–112, 2005.

[25] Adam R. Klivans, Homin K. Lee, and Andrew Wan. Mansour's conjecture is true for random DNF formulas. In *COLT 2010 - The 23rd Conference on Learning Theory, Haifa, Israel, June 27-29, 2010*, pages 368–380, 2010.

[26] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, and R. Urbankez. Comparing the bit-map and block-map decoding thresholds of Reed-Muller codes on BMS channels. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 1755–1759, July 2016.

[27] Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D Pfister, Eren Şaşolu, and Rüdiger L Urbanke. Reed–muller codes achieve capacity on erasure channels. *IEEE Transactions on information theory*, 63(7):4298–4316, 2017.

[28] Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the Fourier spectrum. *SIAM J. Comput.*, 22(6):1331–1348, 1993.

[29] Yishay Mansour. Learning boolean functions via the fourier transform. In *Theoretical advances in neural computation and learning*, pages 391–424. Springer, 1994.

[30] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[31] Ryan O'Donnell and Li-Yang Tan. A composition theorem for the Fourier Entropy-Influence Conjecture. In *ICALP 2013*, pages 780–791, 2013.

[32] Ryan O'Donnell, John Wright, and Yuan Zhou. The Fourier Entropy-Influence Conjecture for certain classes of boolean functions. In *ICALP 2011*, pages 330–341, 2011.

[33] Guy Shalev. On the Fourier Entropy Influence conjecture for extremal classes. *arXiv preprint arXiv:1806.03646*, 2018.

[34] Andrew Wan, John Wright, and Chenggang Wu. Decision trees, protocols and the entropy-influence conjecture. In *ITCS'14*, pages 67–80, 2014.