Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 52 (2013)

Upper Bounds on Fourier Entropy

Sourav Chakraborty*[1], Raghav Kulkarni†[2], Satya V. Lokam‡[3], and Nitin Saurabh§[4]

[1]Chennai Mathematical Institute, Chennai, India
[2]Centre for Quantum Technologies, Singapore
[3]Microsoft Research India, Bangalore, India
[4]Institute of Mathematical Sciences, Chennai, India

Abstract

Given a function $f : \{0,1\}^n \to \mathbb{R}$, its *Fourier Entropy* is defined to be $-\sum_S \widehat{f}^2(S) \log \widehat{f}^2(S)$, where $\widehat{f}$ denotes the Fourier transform of f. This quantity arises in a number of applications, especially in the study of Boolean functions. An outstanding open question is a conjecture of Friedgut and Kalai (1996), called Fourier Entropy Influence (FEI) Conjecture, asserting that the Fourier Entropy of any Boolean function f is bounded above, up to a constant factor, by the total influence (= average sensitivity) of f.

In this paper we give several upper bounds on the Fourier Entropy of Boolean as well as real valued functions. We give a general bound involving the $(1+\delta)$-th moment of $|S|$ w.r.t. the distribution $\widehat{f}^2(S)$; the FEI conjecture needs the first moment of $|S|$. A variant of this bound uses the first and second moments of sensitivities (average sensitivity being the first moment). We also give upper bounds on the Fourier Entropy of Boolean functions in terms of several complexity measures that are known to be bigger than the influence. These complexity measures include, among others, the logarithm of the number of leaves and the average depth of a decision tree. Finally, we show that the FEI Conjecture holds for two special classes of functions, namely linear threshold functions and read-once formulas.

1 Introduction

Fourier transforms are extensively used in a number of fields such as engineering, physics, and computer science. Within theoretical computer science, Fourier analysis of Boolean functions evolved into one of the most useful and versatile tools; see the book [19] for a comprehensive survey of this area and pointers to literature on this subject. In particular, it plays an important role in numerous results in complexity theory, learning theory, social choice, inapproximability, metric

*sourav@cmi.ac.in
†kulraghav@gmail.com
‡satya@microsoft.com
§nitin@imsc.res.in

ISSN 1433-8092

spaces, etc. If $\widehat{f}$ denotes the Fourier transform of a Boolean function f, then $\sum_{S\subseteq[n]} \widehat{f}^2(S) = 1$ and hence we can define an entropy of the distribution given by $\widehat{f}^2(S)$:

$$\mathbb{H}(f) := \sum_{S\subseteq[n]} \widehat{f}^2(S) \log \frac{1}{\widehat{f}^2(S)}. \tag{1}$$

The Fourier Entropy-Influence (FEI) Conjecture, made by Friedgut and Kalai [9] in 1996, states that for every Boolean function, its Fourier entropy is bounded above by its total influence.

Fourier Entropy-Influence Conjecture: There exist a universal constant C such that for all $f : \{0,1\}^n \to \{+1,-1\}$,

$$\mathbb{H}(f) \leqslant C \cdot \mathsf{Inf}(f), \tag{2}$$

where $\mathsf{Inf}(f)$ is the total influence of f which is the same as the average sensitivity $\mathsf{as}(f)$ of f. The latter quantity may be intuitively viewed as the expected number of coordinates of an input which, when flipped, will cause the value of f to be changed, where the expectation is w.r.t. the uniform distribution on the input assignments of f. Thus, the conjecture intuitively asserts that if the Fourier coefficients of a Boolean function are "smeared out," then its influence must be large, i.e., at a typical input, the value of f changes in several different directions.

1.1 Motivation

Resolving the FEI conjecture is one of the most important open problems in the Fourier analysis of Boolean functions. The original motivation for the conjecture in [9] stems from a study of threshold phenomena in random graphs.

The FEI Conjecture has numerous applications. It implies a variant of *Mansour's Conjecture* [16] stating that for a Boolean function computable by a DNF formula with m terms, most of its Fourier mass is concentrated on $\mathsf{poly}(m)$-many coefficients. A proof of Mansour's conjecture would imply an efficient and accurate *agnostic* learning algorithm for DNF's [10] answering a major open question in computational learning theory.

The FEI conjecture implies the seminal KKL-result [11, 21] that for every Boolean function that is 1 with a constant probability, there is always an input variable with influence at least $\Omega(\log n/n)$.

The FEI conjecture also implies that for any n-vertex graph property, the influence is at least $c(\log n)^2$. The best known lower bound, by Bourgain and Kalai [4], is $\Omega((\log n)^{2-\epsilon})$, for any $\epsilon > 0$.

See [12], [20] and [13] for a detailed explanation on these and other consequences of the conjecture.

1.2 Prior Work

The first advance on FEI conjecture was made by Klivans, Lee, and Wan in [13] showing that the conjecture holds for random DNFs. In [20], O'Donnell, Wright, and Zhou proved that the conjecture holds for symmetric functions and more generally for any d-part symmetric functions for constant d. They also proved the conjecture for functions computable by read-once decision trees. Among other recent efforts on the FEI conjecture, Keller, Mossel, and Schlank [12] generalize the conjecture to biased product measures on the Boolean cube and prove a variant of the conjecture for function with extremely low Fourier weight on the high levels. It is also relatively easy to show that the FEI conjecture holds for a random Boolean function, e.g., see Das, Pal, and Visavalia [6] for a proof. By direct calculation, one can verify the conjecture for simple functions like AND, OR, Majority, Tribes etc.

1.3 Our results

We report here various upper bounds on Fourier entropy that may be viewed as progress toward the FEI conjecture.

General bounds: Recall [11] the well-known identity $\mathsf{as}(f) = \mathsf{Inf}(f) = \sum_S |S|\, \widehat{f}^2(S)$ that relates the influence $\mathsf{Inf}(f)$ or average sensitivity $\mathsf{as}(f)$ to $\hat{f}$. Hence, the FEI conjecture (2) states that there is an absolute constant C such that for all Boolean f,

$$\mathbb{H}(f) \leqslant C \cdot \sum_S |S|\, \widehat{f}^2(S). \tag{3}$$

We prove here that for all δ, $0 < \delta \leqslant 1$, and for all f with $\sum_S \widehat{f}^2(S) = 1$, and hence for Boolean f in particular,

$$\mathbb{H}(f) \leqslant \sum_S |S|^{1+\delta}\, \widehat{f}^2(S) \;+\; (\log n)^{O(1/\delta)}. \tag{4}$$

From this, we can also derive

$$\mathbb{H}(f) \leqslant \mathsf{as}(f)^{1-\delta} \cdot \mathsf{as}_2(f)^{\delta} \;+\; (\log n)^{O(1/\delta)}, \tag{5}$$

where $\mathsf{as}_2(f) := \sum_S |S|^2\, \widehat{f}^2(S)$.

Using the "tensorizability" property of the FEI conjecture, O'Donnell et al [20] observe that the FEI conjecture is equivalent to showing that for all $f : \{0,1\}^n \to \{+1,-1\}$,

$$\mathbb{H}(f) \leqslant C \sum_S |S|\, \widehat{f}^2(S) \;+\; o(n). \tag{6}$$

It is worth contrasting (4) and (5) with (6).

It is important to note that (4) holds for *arbitrary*, i.e., even non-Boolean, f such that (without loss of generality) $\sum_S \widehat{f}^2(S) = 1$. On the other hand, there are examples of non-Boolean f for which the FEI conjecture (3) is *false*. From (5), one can show that for *all* f, $\mathbb{H}(f) = O(\mathsf{as}(f) \log n)$. Hence proving the FEI conjecture should involve removing the "extra" log factor while exploiting the Boolean nature of f.

Upper bounds by Complexity Measures: The $\mathsf{Inf}(f)$ of a Boolean function f is used to derive lower bounds on a number of complexity parameters of f such as the number of leaves or the average depth of a decision tree computing f. Hence a natural weakening of the FEI conjecture is to prove upper bounds on the Fourier entropy in terms of such complexity measures of Boolean functions. By a relatively easy argument, we show that

$$\mathbb{H}(f) = O(\log L(f)), \tag{7}$$

where $L(f)$ denotes the minimum number of leaves in a decision tree that computes f. If $\mathsf{DNF}(f)$ denotes the minimum number of terms in a DNF for the function f, note that $\mathsf{DNF}(f) \leqslant L(f)$). Thus improving (7) with $O(\log \mathsf{DNF}(f))$ on the right hand side would resolve *Mansour's conjecture*. We note that (7) also holds when the queries made by the decision tree involve parities or conjunctions of subsets of variables. It also holds when $L(f)$ is generalized to the number of subcubes in a *subcube partition* that represents f. Note that for a Boolean function

$$\mathsf{Inf}(f) \leqslant \log(L_c(f)) \leqslant \log(L(f)) \leqslant D(f),$$

where $L_c(f)$ is number of subcubes in a *subcube partition* that represents f and $D(f)$ is the minimum depth of a decision tree computing f.

We also prove the following strengthening of (7):

$$\mathbb{H}(f) \leqslant 2\,\bar{d}(f), \tag{8}$$

where $\bar{d}(f)$ denotes the minimum *average depth* of a decision tree computing f (observe that $\bar{d}(f) \leqslant \log(L(f))$). Note that the average depth of a decision tree is also a kind of entropy: it is given by the distribution induced on the leaves of a decision tree when an input is drawn uniformly at random. Thus (8) relates the two kinds of entropy up to a constant factor.

FEI inequality for some Special Classes of Boolean functions: Finally, we prove that the FEI conjecture holds for two special classes of Boolean functions:

- Linear Threshold Functions (LTF's), i.e., functions f such that $f(x) = \mathsf{sign}(w_0 + w_1x_1 + \cdots + w_nx_n)$ for $w_i \in \mathbb{R}$, and
- Read-Once Formulas, i.e., functions computable by a tree with AND and OR gates at internal nodes and each variable occurring *at most once* at the leaves.

Prior to our result for LTF's, FEI is known to be true for unweighted threshold functions, i.e., when $f(x) = \mathsf{sign}(x_1 + \cdots + x_n - \theta)$ for some integer $\theta \in [0..n]$. This is a corollary of the result from [20] that the FEI holds for all symmetric Boolean functions. Our proof for LTF's makes use of an upper bound on the level-k Fourier mass of a Boolean function due to Benjamini et al. [2] (see also Talagrand [22] and O'Donnell's lecture notes [17]) and a recent lower bound of $\frac{1}{2} + c$, for an absolute positive constant c, due to De et al. [7] on the level-$\leqslant 1$ mass of any LTF (also known as the Gotsman-Linial or O'Donnell constant). The fact that this bound is strictly larger than $1/2$ by an absolute constant is critical for our application.

O'Donnell et al [20] also prove that the FEI holds for read-once *decision trees*. Our result for read-once formulas is a strict generalization of their result. For instance, the tribes function is computable by read-once formulas but not by read-once decision trees. Our proof for read-once formulas is a consequence of a kind of tensorizability for $\{0,1\}$-valued Boolean functions. In particular, we show that an inequality similar to the FEI inequality is preserved when functions depending on *disjoint* sets of variables are combined by AND and OR operators.

2 Preliminaries

We recall here some basic facts of Fourier analysis. Consider the space of all functions from $\{0,1\}^n$ to $\mathbb{R}$, endowed with the inner product $\langle f, g\rangle = 2^{-n}\sum_{x\in\{0,1\}^n} f(x)g(x)$. The character functions $\chi_S(x) := (-1)^{\sum_{i\in S} x_i}$ for $S \subseteq [n]$ form an orthonormal basis for this space of functions w.r.t. the above inner product. Thus, every function $f : \{0,1\}^n \longrightarrow \mathbb{R}$ of n boolean variables has the *unique Fourier* expansion:

$$f(x) = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S(x).$$

The vector $\widehat{f} = (\widehat{f}(S))_{S\subseteq[n]}$ is called the Fourier transform of the function f. The Fourier coefficient $\widehat{f}(S)$ of f at S is then given by

$$\widehat{f}(S) = 2^{-n} \sum_{x\in\{0,1\}^n} f(x)\chi_S(x).$$

The norm of a function f is defined to be $\|f\| = \sqrt{\langle f, f\rangle}$. Orthonormality of $\{\chi_S\}$ implies the *Parseval's identity:* $\|f\|^2 = \sum_S \widehat{f}^2(S)$.

We consider Boolean functions with range $\{-1,+1\}$. For an $f : \{0,1\}^n \to \{-1,+1\}$, $\|f\|$ is clearly 1 and hence Parsevals' identity shows that for Boolean functions $\sum_S \widehat{f}^2(S) = 1$. This implies that squared Fourier coefficients can be thought of as a probability distribution and the notion of Fourier entropy (1) is well-defined.

The *influence of* f *in the* i*-th direction*, denoted $\mathrm{Inf}_i(f)$ is the fraction of inputs at which the value of f gets flipped if we flip the i-th bit:

$$\mathrm{Inf}_i(f) = 2^{-n}|\{x \in \{0,1\}^n : f(x) \neq f(x \oplus e_i)\}|,$$

where $x \oplus e_i$ is obtained from x by flipping the ith bit of x.

The (total) *influence* of f is defined to by $\mathrm{Inf}(f)$, is $\sum_{i=1}^n \mathrm{Inf}_i(f)$. The influence of i on f can be shown, e.g., [11], to be

$$\mathrm{Inf}_i(f) = \sum_{S\ni i} \widehat{f}(S)^2$$

and hence it follows that $\mathrm{Inf}(f) = \sum_{S\subseteq[n]} |S|\widehat{f}(S)^2$.

For $x \in \{0,1\}^n$, the *sensitivity of* f *at* x, denoted $s_f(x)$, is defined to be $s_f(x) := |\{i : f(x) \neq f(x\oplus e_i), 1 \leqslant i \leqslant n\}|$, i.e., the number of coordinates of x, which when flipped, will flip the value of f. The (maximum) *sensitivity of the function* f, denoted $s(f)$ is defined to be the largest sensitivity of f at x over all $x \in \{0,1\}^n$: $s(f) := \max\{s_f(x) : x \in \{0,1\}^n\}$. The *average sensitivity of* f, denoted $\mathrm{as}(f)$, is defined to be $\mathrm{as}(f) := 2^{-n}\sum_{x\in\{0,1\}^n} s_f(x)$. It is easy to see that $\mathrm{Inf}(f) = \mathrm{as}(f)$ and hence we also have $\mathrm{as}(f) = \sum_{S\subseteq[n]} |S|\widehat{f}(S)^2$.

3 A bound on the entropy of general functions

Theorem 3.1. *If* $f = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S$ *is a real-valued function on the domain* $\{0,1\}^n$ *such that* $\sum_S |\hat{f}(S)^2| = 1$ *then for any* $\delta > 0$

$$\sum_{S\subseteq[n]} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) = \sum_S |S|^{1+\delta}\,\hat{f}(S)^2 + 2\log_{1+\delta} n + 2(2\log n)^{(1+\delta)/\delta}(\log n)^2.$$

3.1 Proof of Theorem 3.1

Our proof strategy is as follows: We partition the Fourier coefficients into suitable parts and then upper bound each part. We start with suitably chosen sets $A_0, B_0 \subseteq 2^{[n]}$ and then inductively construct the sets $A_1, B_1, \ldots, A_k, B_k$. The A_i's represent the new Fourier coefficients whose total entropy we are able to upper bound. The B_i's represent the Fourier coefficients that are *not yet accounted for*. Our construction yields that as k increases B_k only consists of those $\hat{f}(S)$ for which $|S| < \psi(k,n,\delta)$, where ψ is a suitable function of k, n and δ. Finally an appropriate choice of k gives us the desired inequality. We start by describing the A_i and B_i.

Let A_0 be the be the set of all $S \subseteq [n]$ for which $|S|^{1+\delta}$ is at least $\log\left(\frac{1}{\hat{f}(S)^2}\right)$. That is:

$$A_0 := \{S \mid \hat{f}(S)^2 \geqslant 1/2^{|S|^{1+\delta}}\}.$$

Clearly,

$$\sum_{S\in A_0} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) \leqslant \sum_{S\in A_0} |S|^{1+\delta}\hat{f}(S)^2. \tag{9}$$

Now, let A_1 be all the $S \subseteq [n]$ for which $|\hat{f}(S)| \leqslant 2^{-n}$. The following lemma helps to upper bound the contribution of sets from A_1 to the entropy.

Lemma 3.2. *For any* t*, let* $\mathcal{T} \subseteq \{S \mid |\hat{f}(S)| \leqslant 1/t\}$ *and* $|\mathcal{T}| \leqslant t$ *then* $\sum_{S\in\mathcal{T}} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) \leqslant 2$. *Also for any* k *we have*

$$\sum_{S:|S|\leqslant k} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) \leqslant 2 + 2k\log n.$$

Proof. Since $|\hat{f}(S)| \leqslant 1/t$ for every $S \in \mathcal{T}$,

$$\sum_{S\in\mathcal{T}} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) \leqslant \frac{2}{t}\sum_{S\in\mathcal{T}} |\hat{f}(S)| \log\left(\frac{1}{|\hat{f}(S)|}\right) \leqslant 2,$$

where the last inequality follows from the fact that $|\hat{f}(S)|\log(1/|\hat{f}(S)|) < 1$ (because $x\log(1/x) < 1$ for all $0 \leqslant x < 1$) and $|\mathcal{T}| \leqslant t$.

Now for the second inequality note that $\{S \mid |S| \leqslant k\}$ is less than n^k. Let $S_k := \{S | \hat{f}(S) < 1/n^k\}$. From the above inequality $\sum_{S\in S_k} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) \leqslant 2$, Now for all S such that $|S| \leqslant k$ and $S \notin S_k$ then $\log(1/|\hat{f}(S)|) < k\log n$ and hence

$$\sum_{S:|S|\leqslant k \text{ and } S\notin S_k} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) \leqslant 2k\log n.$$

$\square$

Thus by Lemma 3.2 we have,

$$\sum_{S\in A_1} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) \leqslant 2. \tag{10}$$

Let $B_1 = \{0,1\}^n \backslash (A_0 \cup A_1)$. By the definition of A_0 and A_1, $B_1 \subseteq \left\{S \mid \frac{1}{2^{2n}} \leqslant \hat{f}(S)^2 \leqslant \frac{1}{2^{|S|^{1+\delta}}}\right\}$. Thus $B_1 \subseteq \{S \mid |S| \leqslant (2n)^{1/1+\delta}\}$ and hence $|B_1| \leqslant \binom{n}{(2n)^{1+\delta}} < n^{(2n)^{1+\delta}}$. Hence by Lemma 3.2 we have a $o(n)$ contribution from the total entropy of B_1 and thus from Equation 9 and 10 we obtain

$$\sum_{S\subseteq[n]} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) = \sum_{S\in A_0} |S|^{1+\delta}\hat{f}(S)^2 + o(n).$$

Now we proceed to sharpen the $o(n)$ term. We repeat the above step via an inductive construction described as follows: Suppose we have constructed B_k and let r_k be the smallest such that $|B_k| \leqslant n^{r_k}$. Define $A_{k+1} := \{S \in B_k \mid \hat{f}(S)^2 \leqslant 1/n^{2r_k}\}$, and $B_{k+1} := B_k \backslash A_{k+1}$. So,

$$B_{k+1} \subseteq \left\{S \mid \frac{1}{n^{2r_k}} \leqslant \hat{f}(S)^2 \leqslant \frac{1}{2^{|S|^{1+\delta}}}\right\}.$$

Thus $B_{k+1} \subseteq \{S \mid |S| \leqslant (2r_k \log n)^{1/1+\delta}\}$ and we obtain the r_{k+1} which is smallest such that $|B_{k+1}| \leqslant n^{r_{k+1}}$. Thus $r_{k+1} \leqslant (2r_k \log n)^{1/1+\delta}$. We know that $r_1 \leqslant n$. For $k = \log_{1+\delta} n$ we have $r_k \leqslant (2\log n)^{(1+\delta)/\delta}$. Note that for all $k \geqslant 0$ since $A_{k+1} \subseteq B_k$ and $|B_k| \leqslant n^{r_k}$ thus by the definition of A_{k+1} and by Lemma 3.2

$$\sum_{S\in A_{k+1}} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) \leqslant 2.$$

Thus each A_k contributes at most 2 to the entropy. The only Fourier coefficients that we have not accounted for after k^{th} step are the ones in B_{k+1} and those are with $|S| \leqslant (2r_k \log n)^{1/(1+\delta)} \leqslant 2(2\log n)^{(1+\delta)/\delta} \log n$. Hence from Lemma 3.2 we have

$$\sum_{S\subseteq[n]} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) \leqslant \sum_{S\subseteq[n]} |S|^{1+\delta}\hat{f}(S)^2 + 2\log_{1+\delta} n + 2(2\log n)^{(1+\delta)/\delta}(\log n)^2.$$

This completes the proof of Theorem 3.1. $\square$

3.2 Corollaries to Theorem 3.1

By choosing $\delta = \frac{\log\log n}{\log n}$, we obtain

Corollary 3.3. *For any function $f : \{0,1\}^n \to \mathbb{R}$ such that $\sum_S |\hat{f}(S)^2| = 1$ then*

$$\sum_{S\subseteq[n]} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) = O(as(f)\log n) + o(n).$$

Corollary 3.4. *If* $f = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S$ *is a real-valued function on the domain* $\{0,1\}^n$ *such that* $\sum_S |\widehat{f}(S)^2| = 1$ *then for any* $\delta > 0$

$$\sum_{S\subseteq[n]} \widehat{f}(S)^2 \log\left(\frac{1}{\widehat{f}(S)^2}\right) = as(f)^{1-\delta} as_2(f)^{\delta} + 2\log_{1+\delta} n + 2(2\log n)^{(1+\delta)/\delta}(\log n)^2,$$

where $as_2(f) := \sum_S |S|^2\, \widehat{f}(S)^2$.

The proof of Corollary 3.4 follows from Lemma 3.6 below. But for proving Lemma 3.6 we need Lemma 3.5. We prove the lemmas below.

For a Boolean function $f : \{0,1\}^n \to \{1,-1\}$ we say that the i^{th} bit f is *sensitive* on input x if $f(x \oplus e_i) \neq f(x)$, where $e_i \in \{0,1\}^n$ is a vector whose all co-ordinates except the i^{th} are 0 as x, and the i^{th} bit is 1. The sensitivity of f on input x, denoted by $s(f,x)$ is the total number of sensitive bits of f on x. The sensitivity of f, denoted by $s(f)$, is the maximum over all x of $s(f,x)$. The average sensitivity denoted by $as(f) := \sum_x s(f,x)/2^n$. Kahn-Kalai-Linial [11] proved that

$$as(f) = \sum_S \widehat{f}(S)^2|S| = \mathrm{Inf}(f).$$

For $f : \{0,1\}^n \to \mathbb{R}$, let $as_2(f) := \sum_S |S|^2\widehat{f}(S)^2$. The following lemma gives an identity for $as_2(f)$ in terms of the sensitivity of f. We present here a proof due to Alex Samorodnitsky (personal communication).

Lemma 3.5. *For* $f : \{-1,1\}^n \to \{-1,1\}$,

$$\frac{1}{2^n}\sum_x s(f,x)^2 = \sum_{S\subseteq[n]} |S|^2\widehat{f}(S)^2 = as_2(f).$$

Proof. Consider the following function $L : \{0,1\}^n \to \mathbb{R}$

$$L(x) = \begin{cases} n & \text{for} \quad |x| = 0 \\ -1 & \text{for} \quad |x| = 1 \\ 0 & \text{for} \quad |x| > 1 \end{cases}$$

Let $L * f(x) := \frac{1}{2^n}\sum_z L(x \oplus z)f(z)$. Note that:

$$(L * f)(x) = \frac{2s(f,x)f(x)}{2^n}.$$

Using Parseval's Identity we obtain:

$$\frac{1}{2^n}\sum_x (2s(f,x)/2^n)^2 = \sum_{S\subseteq[n]} \widehat{(L * f)}(S)^2 = \sum_{S\subseteq[n]} \hat{L}(S)^2\hat{f}(S)^2.$$

Since for any $S \subseteq [n]$, $\hat{L}(S) = 2|S|/2^n$ so we obtain the equality that

$$\frac{1}{2^n}\sum_x 4s(f,x)^2 = \sum_{S\subseteq[n]} 4|S|^2\hat{f}(S)^2.$$

□

Lemma 3.6. *For all* $0 \leqslant \delta \leqslant 1$,

$$\mathrm{as}(f)^{1-\delta}\mathrm{as}_2(f)^{\delta} \geqslant \sum_{S \subseteq [n]} |S|^{1+\delta}\widehat{f}(S)^2.$$

Proof. We treat $\widehat{f}(S)^2$ as the probability associated to the set S and use the following version of the Cauchy-Schwartz inequality: for any two random variables $X, Y : \Omega \to \mathbb{R}$, we have: $\sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)} \geqslant \mathbb{E}(XY)$. Choosing $X(S) = \sqrt{|S|}$ and $Y(S) = |S|$ immediately yields the desired inequality for the value of $\delta = \frac{1}{2}$ in light of Lemma 3.5.

In fact, we can show the following: if the desired inequality holds for $\delta = \alpha$ and $\delta = \beta$ then the inequality must also hold for $\delta = \frac{\alpha+\beta}{2}$. To show this, one may apply the Cauchy-Schwartz inequality with $X(S) = |S|^{(1+\alpha)/2}$ and $Y(S) = |S|^{(1+\beta)/2}$.

Hence, by continuity, the desired inequality holds for any $\delta \in [0, 1]$. $\square$

4 Bounding Entropy using Complexity Measures

In this section, we prove upper bounds on Fourier entropy in terms of some complexity parameters associated to decision trees and subcube partitions.

4.1 *via* L_1-norm : Decision Trees

Lemma 4.1. *Let* $f : \{0,1\}^n \to \mathbb{R}$ *be such that* $\sum_S \widehat{f}(S)^2 \leqslant 1$. *Let* $L_1(f) := \sum_S |\widehat{f}(S)|$ *be the* L_1*-norm of the Fourier transform of* f. *Assume, further, that* $L_1(f) \geqslant 1$. *Then,* $\mathbb{H}(f) \leqslant 4 \log L_1(f) + 9$.

Proof. Let $L := L_1(f)$. Let $\theta := 1/(16L^2)$. Let $\mathcal{G} := \{S : |\widehat{f}(S)| \geqslant \theta\}$. Note that for $x \geqslant 16$, $\log x \leqslant \sqrt{x}$. We thus have for $S \notin \mathcal{G}$, $\log \frac{1}{|\widehat{f}(S)|} \leqslant \frac{1}{\sqrt{|\widehat{f}(S)|}}$.

$$\begin{aligned}
\mathbb{H}(f) = \sum_S \widehat{f}^2(S) \log \frac{1}{\widehat{f}^2(S)} &\leqslant \sum_{S \in \mathcal{G}} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2} + 2 \sum_{S \notin \mathcal{G}} \widehat{f}(S)^2 \frac{1}{\sqrt{|\widehat{f}(S)|}} \\
&\leqslant \log \frac{1}{\theta^2} \sum_{S \in \mathcal{G}} \widehat{f}(S)^2 + 2 \max_{S \notin \mathcal{G}} \sqrt{|\widehat{f}(S)|} \sum_{S \notin \mathcal{G}} |\widehat{f}(S)| \\
&\leqslant \log(256 L^4) + 2 \cdot \frac{1}{4L} \cdot L \leqslant 4 \log L + 9.
\end{aligned}$$

$\square$

It is well-known and easy to prove[1] that the L_1-norm of a function computed by a decision tree is at most the number of leaves in that tree.

Lemma 4.2 ([14, 15]). *Let* $\ell(f)$ *be the minimum number of leaves in a decision tree computing the Boolean function* f. *Then* $L_1(f) := \sum_S |\widehat{f}(S)| \leqslant \ell(f)$.

In fact, even if we allow the queries at each internal node of a decision tree to be parities or conjunctions of subsets of variables (or more generally any functions with bounded L_1*-norm), then also we have* $L_1(f) = O(\ell(f))$.

[1] See also Lemma 4.8

Corollary 4.3. *Let $\ell(f)$ be the number of leaves in a decision tree computing a Boolean function f. Then $\mathbb{H}(f) = O(\log \ell(f))$.*

The same bound as above holds also for decision trees that query parities and conjunctions of variables at their internal nodes.

The decision tree *depth* $D(f)$ of a function is the minimum depth (length of a longest root-to-leaf path) of a decision tree computing f and that the *degree* $\deg(f)$ is the degree of the (unique) multilinear polynomial over $\mathbb{R}$ that represents f. It is easy to see that $\log(L_1(f)) \leqslant \deg(f) \leqslant D(f)$. Thus, we immediately have

Corollary 4.4. *For every Boolean function f, $\mathbb{H}(f) = O(D(f))$ and $\mathbb{H}(f) = O(\deg(f))$.*

Remark 4.1. *A natural question to ask is how important Boolean-ness of functions is in entropy upper bounds. While Lemma 4.1 holds for real-valued functions as well, we note that the Corollaries 4.3 and 4.4 hold* only for Boolean-valued *functions. In fact, we give examples below to show that these corollaries fail for non-Boolean functions.*

A decision tree for a non-Boolean, say $\mathbb{R}$-valued, function f can be defined by a natural generalization of the one for a Boolean-valued function. It queries the (Boolean) input variables as in the usual decision tree, but produces a value in $\mathbb{R}$ at each leaf. It must guarantee that on all inputs that reach a leaf the function value must be constant and equal to the value produced at that leaf.

Our next example shows that Fourier entropy cannot be upper bounded by log(number of leaves) for non-Boolean f in contrast to Corollary 4.3 for Boolean functions. In fact, there is an exponential gap:

Lemma 4.5. *There exists a function $f : \{0,1\}^n \to \mathbb{R}$ satisfying $\sum_S \widehat{f}^2(S) = 1$ such that*

$$\sum_{S \subseteq [n]} \widehat{f}(S)^2 \log\left(\frac{1}{\widehat{f}(S)^2}\right) = \Omega(n), \quad \textit{but} \quad \log \ell(f) = O(\log n).$$

Proof. Consider the following function:

$$f(x) = \sqrt{\frac{2^{d(x)}}{n}},$$

where $d(x)$ is the first index in x that is 1. Note that this function has a decision tree same as the OR function and thus have only $n+1$ leaves. Now to see that $\sum_{S \subseteq [n]} \widehat{f}(S)^2 = 1$ consider the following:

$$\sum_x f(x)^2 = \sum_{i \in [n]} \sum_{x : d(x) = i} f(x)^2 = \sum_{i \in [n]} 2^{n-i} \frac{2^i}{n} = 2^n,$$

and thus from Parseval's identity we have $\sum_{S \subseteq [n]} \widehat{f}(S)^2 = 1$.

It is easy to check that for any set $S \subseteq [n]$ if k is the largest index in S then

$$\widehat{f}(S) = \frac{1}{\sqrt{n} 2^n} \left(2^{n-k} \sqrt{2}^k - \frac{\sqrt{2}^n - \sqrt{2}^k}{\sqrt{2} - 1} \right) \approx \frac{1}{\sqrt{n 2^k}}.$$

And from this it follows that the entropy for the fourier coefficient squares is around $n/2 + \log n$ whereas $\log(\ell) = \log(n)$. ∎

Our next example shows that Fourier entropy can be logarithmically larger than the degree for non-Boolean functions in contrast to Corollary 4.4 for Boolean functions.

Lemma 4.6. *There exists a function* $f : \{0,1\}^n \to \mathbb{R}$ *of degree* d *satisfying* $\sum_S |\hat{f}(S)^2| = 1$ *such that*

$$\sum_{S\subseteq[n]} \hat{f}(S)^2 \log\left(\frac{1}{\hat{f}(S)^2}\right) = \Omega(d \log n).$$

Proof. Consider the following function $f = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S$, where $\hat{f}(S) = 1/\sqrt{\binom{n}{2}}$ if $|S| = 2$, and $\hat{f}(S) = 0$ otherwise. It is easy to see that the $\mathbb{H}(f) = \log\binom{n}{2}$, whereas $\mathsf{Inf}(f) = \sum_{S\subseteq[n]} |S|\hat{f}(S)^2 = 2$.

So now if we put uniform weights on k-sized sets, that is, $\hat{f}(S) = 1/\sqrt{\binom{n}{k}}$ if $|S| = k$ and $\hat{f}(S) = 0$ if $|S| \neq k$, we will get $\mathsf{Inf}(f) = k$ and $\mathbb{H}(f) = \log\binom{n}{k} \geqslant k\log n - k\log k$. Choosing $k = \sqrt{n}$, we will have $\mathbb{H}(f) = \Omega(\sqrt{n}\log n)$ and $\mathsf{Inf}(f) = \sqrt{n}$. Since the degree of the function is $d = \sqrt{n}$, we get $\mathbb{H}(f) = \Omega(d \cdot \log n)$. ☐

4.2 *via* Concentration : Subcube Partitions

Note that a decision tree computing a Boolean function f induces a partition of the cube $\{0,1\}^n$ into monochromatic subcubes, i.e., f has the same value on all points in a given subcube, with one subcube corresponding to each leaf. But there exist monochromatic subcube partitions that are not induced by any decision tree. We can generalize Corollary 4.3 to subcube partitions.

Our goal in presenting the generalization to subcube partitions is to illustrate a different approach. The approach uses the concentration property of the Fourier transform and uses a general, potentially powerful, technique. One way to do this is to use a result due to Bourgain and Kalai (Theorem 3.2 in [12]). However, we give a more direct proof for the special case of subcube partitions.

Definition 4.7. *A* subcube C *of the cube* $B_n := \{0,1\}^n$ *is given by a mapping (partial assignment)* $\alpha : [n] \to \{-1,+1,*\}$ *and is defined to be the set of all vectors in* B_n *that agree with* α *on coordinates* fixed, *i.e., assigned a non-$*$ value, by* α*:* $C := C_\alpha := \{x \in B_n \ : \ \alpha(i) \neq * \implies x_i = \alpha(i)\}$. *We use* $A := \{i \in [n] : \alpha(i) \neq *\}$ *to denote the set of fixed coordinates of* α *and denote the cube* C *also by the pair* (A, α).

For a function $f : \{0,1\}^n \to \{+1,-1\}$*, a partition* $\mathcal{C} = \{C_1, \ldots, C_m\}$ *of* B_n *into subcubes* C_i *such that* f *is constant on each* C_i *is called a* (monochromatic) subcube partition *w.r.t.* f*. If* $\mathcal{C}$ *is a subcube partition monochromatic w.r.t.* f*, we also say* $\mathcal{C}$ computes f.

We denote by $L_c(f)$ *the minimum number of subcubes in a subcube partition that computes* f.

Suppose now f is computed by the subcube partition $\mathcal{C} = \{C_1, \ldots, C_L\}$, where $C_i = (A_i, \alpha_i)$. Let $\phi_i : \{0,1\}^n \to \{0,1\}$ be the characteristic function of the subcube C_i : $\phi_i(x) = 1$ if $x \in C_i$ and $\phi_i(x) = 0$ otherwise. Let $\beta_i \in \{-1,+1\}$ be the value of f on C_i. Then, clearly

$$f(x) = \sum_{i=1}^{L} \beta_i \phi_i(x).$$

By linearity of the Fourier transform, it follows that $\hat{f}(S) = \sum_{i=1}^{L} \beta_i \hat{\phi}_i(S)$. A simple calculation shows that, for the characteristic function ϕ of a subcube $C = (A, \alpha)$, the Fourier transform is given by $\hat{\phi}(S) = 2^{-|A|}\chi_S(\alpha)$ if $S \subseteq A$ and $\hat{\phi}(S) = 0$ otherwise. It follows that

$$\hat{f}(S) = \sum_{i\,:\,S\subseteq A_i} 2^{-|A_i|} \cdot \beta_i \chi_S(\alpha_i). \tag{11}$$

In particular, $\hat{f}(S) \neq 0 \implies \exists i\ S \subseteq A_i$.

The following lemma directly follows from the above observations.

Lemma 4.8. *Let* f *be computed by the subcube partition* $\mathcal{C} = \{C_1, \ldots, C_L\}$*, where* $C_i = (A_i, \alpha_i)$*. Then,*

(i) $\sum_S |\hat{f}(S)| \leqslant L$, *and*

(ii) *For any integer* $t \geqslant 0$, $\sum_{|S|\geqslant t} \hat{f}^2(S) \leqslant \sum_{|A_i|\geqslant t} 2^{-|A_i|}$.

Proof. Using (11),

$$\sum_S |\hat{f}(S)| = \sum_S \Big|\sum_{i\,:\,S\subseteq A_i} \beta_i \chi_S(\alpha_i) 2^{-|A_i|}\Big| \leqslant \sum_S \sum_{i\,:\,S\subseteq A_i} 2^{-|A_i|} = \sum_{i=1}^{L} 2^{-|A_i|} \sum_{S\subseteq A_i} 1 = \sum_{i=1}^{L} 2^{-|A_i|} \cdot 2^{|A_i|} = L.$$

By (11), if $|S| \geqslant t$, the contribution to $\hat{f}(S)$ comes from only the C_i such that $|A_i| \geqslant t$. Let $g \equiv \sum_{|A_i|\geqslant t} \beta_i \phi_i$ be the restriction of f to subcubes with codimension $\geqslant t$. It is then clear that

$$\sum_{|S|\geqslant t} \hat{f}^2(S) = \sum_{|S|\geqslant t} \hat{g}^2(S) \leqslant \sum_S \hat{g}^2(S) = 2^{-n} \sum_{|A_i|\geqslant t} |C_i| = \sum_{|A_i|\geqslant t} 2^{-|A_i|}.$$

This proves (ii). ☐

Combining Lemma 4.8(i) and Lemma 4.1, it immediately follows that $\mathbb{H}(f) = O(\log L_c(f))$.

However, we give here a different approach to prove essentially the same result. The approach uses the concentration property of the Fourier transform and illustrates a general, potentially powerful, technique. One way to do this is to combine Lemma 4.8(ii) with the following result due to Bourgain and Kallai:

Theorem 4.9 (Bourgain-Kalai, cited in [12]). *For* $f : \{0,1\}^n \to \{+1, -1\}$ *,suppose that there exist* $c_0 > 0$, $0 < a < 1/2$, *and integer* k *such that for all* t,

$$\sum_{S:|S|>t} \hat{f}^2(S) \leqslant e^{c_0 k} \cdot e^{-at}.$$

Then, for any $\alpha > 1$, *there exists a set* $\mathcal{B}_\alpha$ *such that*

(i) $\log |\mathcal{B}_\alpha| \leqslant C \cdot \alpha k$, *where* C *depends only on* a *and* c_0, *and*

(ii) $\sum_{S\notin\mathcal{B}_\alpha} \hat{f}^2(S) \leqslant n^{-\alpha}$.

However, we gave a more direct proof that nevertheless derives statements analogous to (i) and (ii) of Theorem 4.9, but for the special case of subcube partitions.

Theorem 4.10. *Let* $f : \{0,1\}^n \to \{+1,-1\}$ *be computed by a subcube partition* $\mathcal{C}$ *of size* L. *Then,*

$$\mathbb{H}(f) \leqslant 2\log L(f) + 2\log n + 2.$$

Proof. To bound entropy via concentration, we use the following simple idea. Suppose $\mathcal{E}$ is a subset of Fourier coefficients of a Boolean function f such that $\sum_{S\in\mathcal{E}} \widehat{f}^2(S) = \epsilon$. For a subset of coefficients $\mathcal{B}$, let $\mathbb{H}(\mathcal{B})$ denote the Fourier entropy restricted to that set $\mathcal{B}$, appropriately normalized. Then a simple manipulation shows

$$\sum_S \widehat{f}^2(S)\log\frac{1}{\widehat{f}^2(S)} = (1-\epsilon)\,\mathbb{H}(\overline{\mathcal{E}}) + \epsilon\,\mathbb{H}(\mathcal{E}) + H(\epsilon), \tag{12}$$

where $H(p) := p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}$ is the binary entropy function.

Now, let

$$\mathcal{B}_t := \{S \,:\, \exists\, i\; |A_i| \leqslant t \;\text{ such that }\; S \subseteq A_i\}.$$

Note that if $S \notin \mathcal{B}_t$, then every set A_i that contains S must have size larger than t. Hence, using (??), only sets of size larger than t contribute to such $\widehat{f}(S)$. We now argue as in the proof of Lemma 4.8(ii). Let $g \equiv \sum_{|A_i|>t} \beta_i\phi_i$ be the restriction of f to subcubes with codimension $> t$. It is then clear that

$$\sum_{S\notin\mathcal{B}_t} \widehat{f}^2(S) = \sum_{S\notin\mathcal{B}_t} \widehat{g}^2(S) \leqslant \sum_S \widehat{g}^2(S) = 2^{-n}\sum_{|A_i|>t} |C_i| = \sum_{|A_i|>t} 2^{-|A_i|} < 2^{-t}L. \tag{13}$$

Since $\sum_i 2^{-|A_i|} = 1$, we have that $|\{i \,:\, |A_i| \leqslant t\}| \leqslant 2^t$. Since every $S \in \mathcal{B}_t$ is a subset of some A_i with $|A_i| \leqslant t$, it follows

$$|\mathcal{B}_t| \leqslant \sum_{|A_i|\leqslant t} 2^{|A_i|} \leqslant 2^t \cdot |\{i \,:\, |A_i| \leqslant t\}| \leqslant 2^{2t}. \tag{14}$$

Fix $t := \log(Ln)$.

We can now estimate the Fourier entropy of a subcube partition:

$$\begin{aligned}
\mathbb{H}(f) &= \sum_S \widehat{f}^2(S)\log\frac{1}{\widehat{f}^2(S)} \\
&= (1-1/n)\,\mathbb{H}(\widehat{f}^2(S) \,:\, S\in\mathcal{B}_t) + (1/n)\,\mathbb{H}(\widehat{f}^2(S) \,:\, S\notin\mathcal{B}_t) + H(1/n) \quad \text{using (12) and (13)} \\
&\leqslant (1-1/n)\log|\mathcal{B}_t| + 1/n\cdot n + H(1/n) \;\; \text{using (14)} \\
&\leqslant 2t + 1 + H(1/n) \\
&\leqslant 2\log L + 2\log n + 2.
\end{aligned}$$

∎

4.3 *via* leaf entropy : Average Decision Tree Depth

Let T be a decision tree computing $f : \{0,1\}^n \to \{+1,-1\}$ on variable set $X = \{x_1,\ldots,x_n\}$. If $A_1,\ldots,A_L$ are the sets (with repetitions) of variables queried along the root-to-leaf paths in the tree T, then the average depth (w.r.t. the uniform distribution on inputs) of T is defined to be

$$\bar{d} := \sum_{i=1}^{L} |A_i| 2^{-|A_i|}.$$

Note that the average depth of a decision tree is also a kind of entropy: if each leaf λ_i is chosen with the probability $p_i = 2^{-|A_i|}$ of uniformly chosen random input reaches it, then the entropy of the distribution induced on the λ_i is $\mathbb{H}(\lambda_i) = -\sum_i p_i \log p_i = \sum_i |A_i| 2^{-|A_i|}$. Here, we will show that *the Fourier entropy is at most twice the leaf entropy of a decision tree.*

W.l.o.g., let x_1 be the variable queried by the root node of T and let T_1 and T_2 be the subtrees reached by the branches $x_1 = +1$ and $x_1 = -1$ respectively and let g_1 and g_2 be the corresponding functions computed on variable set $Y = X \setminus \{x_1\}$. Let $\bar{d}$ be the average depth of T and $\bar{d_1}$ and $\bar{d_2}$ be the average depths of T_1 and T_2 respectively. The following straightforward lemma relates the Fourier transform of f to those of it subfunctions g_1 and g_2.

Lemma 4.11. *Let $S \subseteq \{2,\ldots,n\}$.*

(i) $\widehat{f}(S) = (\widehat{g_1}(S) + \widehat{g_2}(S))/2$.

(ii) $\widehat{f}(S \cup \{1\}) = (\widehat{g_1}(S) - \widehat{g_2}(S))/2$.

(iii) $\bar{d} = (\bar{d_1} + \bar{d_2})/2 + 1$.

Proof. Observe that

$$\begin{aligned} f(x_1,x_2,\ldots,x_n) = f(x_1,y) &= \frac{(1+x_1)}{2} g_1(y) + \frac{(1-x_1)}{2} g_2(y) \\ &= \frac{1}{2}(g_1(y) + g_2(y)) + \frac{x_1}{2}(g_1(y) - g_2(y)). \end{aligned}$$

(i) and (ii) now follow by linearity of the Fourier transform.

Let $\{B_i\}_{i=1}^{L_1}$ be the variable sets queried along the root-to-leaf paths in T_1 and similarly let $\{C_i\}_{i=1}^{L_2}$ be the variable sets queried along the root-to-leaf paths in T_2. Then, note that the variable sets $\{A_1\}_{i=1}^{L}$, where $L = L_1 + L_2$, queried along the root-to-leaf paths in T are given by $\{B_i \cup \{x_1\}\}_{i=1}^{L_1} \bigcup \{C_i \cup \{x_1\}\}_{i=1}^{L_2}$. It thus follows that

$$\begin{aligned} \bar{d} &:= \sum_{i=1}^{L} |A_i| 2^{-|A_i|} \\ &= \sum_{1=1}^{L_1} (|B_i| + 1) 2^{-|B_i|-1} \ + \ \sum_{1=1}^{L_2} (|C_i| + 1) 2^{-|C_i|-1} \\ &= \frac{1}{2} \sum_{1=1}^{L_1} |B_i| 2^{-|B_i|} + \frac{1}{2} \sum_{1=1}^{L_1} 2^{-|B_i|} + \frac{1}{2} \sum_{1=1}^{L_2} |C_i| 2^{-|C_i|} + \frac{1}{2} \sum_{1=1}^{L_2} 2^{-|C_i|} \\ &= \frac{1}{2}\bar{d_1} + \frac{1}{2} + \frac{1}{2}\bar{d_2} + \frac{1}{2}, \end{aligned}$$

where the last line follows by applying the definition of average depth to T_1 and noting that $\sum_{1=1}^{L_1} 2^{-|B_i|} = 1$ for the decision tree T_1 and similarly for T_2. This proves (iii). ☐

Remark 4.2. *Note that g_1 and g_2 differ on an input y if and only f is sensitive to x_1 at (x_1, y). In particular, it is obvious that $\frac{1}{4}\|g_1 - g_2\|^2$ is the probability with which f changes value at a random input when x_1 is flipped and hence is $\mathrm{Inf}_f(1)$. Not surprisingly, from (ii), $\sum_{1\in T} \widehat{f}^2(T) = \frac{1}{4}\sum_{S\subseteq Y} \widehat{g_1 - g_2}^2(S) = \frac{1}{4}\|g_1 - g_2\|^2$ and we know that $\mathrm{Inf}_f(1) = \sum_{1\in T} \widehat{f}^2(T)$. It also follows from this that $\frac{1}{4}\|g_1 + g_2\|^2 = \frac{1}{4}\sum_{S\subseteq Y} \widehat{g_1 + g_2}^2(S) = 1 - \mathrm{Inf}_f(1)$.*

We will also need the following technical lemma.

Lemma 4.12. *Let g_1 and g_2 be defined as before Lemma 4.11. Then,*

$$\mathbb{H}(f) \leqslant \frac{1}{2}\mathbb{H}(g_1) + \frac{1}{2}\mathbb{H}(g_2) + 2. \tag{15}$$

Proof. We will use the concavity of the function $x \log \frac{1}{x}$ (for $0 \leqslant x \leqslant 1$)[2] and Lemma 4.11. For simplicity of notation below, let $N' := \{2, \ldots, n\}$.

$$\begin{aligned}
\mathbb{H}(f) &= \sum_{T\subseteq[n]} \widehat{f}^2(T) \log \frac{1}{\widehat{f}^2(T)} \\
&= \sum_{S\subseteq N'} \widehat{f}^2(S) \log \frac{1}{\widehat{f}^2(S)} + \widehat{f}^2(S\cup\{1\}) \log \frac{1}{\widehat{f}^2(S\cup\{1\})} \\
&\leqslant \sum_{S\subseteq N'} (\widehat{f}^2(S) + \widehat{f}^2(S\cup\{1\})) \log \frac{2}{\widehat{f}^2(S) + \widehat{f}^2(S\cup\{1\})} \qquad \text{by concavity of } x\log\tfrac{1}{x}. \\
&= \sum_{S\subseteq N'} \frac{\widehat{g_1}^2(S) + \widehat{g_2}^2(S)}{2} \log \frac{4}{\widehat{g_1}^2(S) + \widehat{g_2}^2(S)} \qquad \text{by Lemma 4.11 (i) and (ii)} \\
&= \frac{1}{2}\sum_{S\subseteq N'} \widehat{g_1}^2(S) \log \frac{1}{\widehat{g_1}^2(S) + \widehat{g_2}^2(S)} + \frac{1}{2}\sum_{S\subseteq N'} \widehat{g_2}^2(S) \log \frac{1}{\widehat{g_1}^2(S) + \widehat{g_2}^2(S)} + \sum_{S\subseteq N'} \widehat{g_1}^2(S) + \widehat{g_2}^2(S) \\
&\leqslant \frac{1}{2}\sum_{S\subseteq N'} \widehat{g_1}^2(S) \log \frac{1}{\widehat{g_1}^2(S)} + \frac{1}{2}\sum_{S\subseteq N'} \widehat{g_2}^2(S) \log \frac{1}{\widehat{g_2}^2(S)} + 2 \\
&\qquad \text{since } \textstyle\sum_{S\subseteq N'} \widehat{g_1}^2(S) = \sum_{S\subseteq N'} \widehat{g_2}^2(S) = 1.
\end{aligned}$$

☐

Let $\bar{d}(f)$ denote the minimum *average* depth of a decision tree computing f.

Theorem 4.13. *For every Boolean function f, $\mathbb{H}(f) \leqslant 2 \cdot \bar{d}(f)$.*

Proof. The proof is by induction on the number of variables of f.

$$\mathbb{H}(f) = \frac{1}{2}\mathbb{H}(g_1) + \frac{1}{2}\mathbb{H}(g_2) + 2 \quad \text{by Lemma 4.12}$$

[2] That is, $x \log \frac{1}{x} + y \log \frac{1}{y} \leqslant (x+y) \log \frac{2}{x+y}$.

$$
\begin{aligned}
&\leqslant \bar{d}_1 + \bar{d}_2 + 2 \quad \text{by induction, } \mathbb{H}(g_i) \leqslant 2\bar{d}_i \text{ for } i = 1, 2 \\
&= 2\bar{d} \quad \text{by Lemma 4.11 (iii)}
\end{aligned}
$$

□

Remark 4.3. *The constant* 2 *in the bound of Theorem 4.13 cannot be replaced by* 1. *Indeed, let* $f(x, y) = x_1y_1 + \cdots + x_{n/2}y_{n/2} \bmod 2$ *be the inner product mod 2 function. Then because* $\widehat{f}^2(S) = 2^{-n}$ *for all* $S \subseteq [n]$, $\mathbb{H}(f) = n$. *On the other hand, it can be shown that* $\bar{d}(f) = \frac{3}{4}n - o(n)$. *Hence, the constant must be at least* 4/3.

5 Entropy-Influence Inequality for some special classes

So far we have established upper bounds on Fourier entropy of general boolean functions. In this section, we will prove the Fourier Entropy-Influence conjecture for balanced linear threshold functions and read-once formulas.

5.1 Linear Threshold Functions

For $f : \{0, 1\}^n \to \{+1, -1\}$, let $W^k[f] := \sum_{S=k} \widehat{f}(S)^2$. Let $\mathbb{II}(f) := \sum_{i=1}^n \mathrm{Inf}_i(f)^2$.

From [20] we have the following bound on entropy:

Lemma 5.1. *[20] Let* $f : \{0, 1\}^n \to \{+1, -1\}$ *be any Boolean function. Then,*

$$
\mathbb{H}(f) \leqslant \sum_{k=0}^{n} W^k[f] \log \binom{n}{k} + \sum_{k=0}^{n} W^k[f] \log \frac{1}{W^k[f]} \leqslant \frac{1}{\ln 2} \mathrm{Inf}[f] + \frac{1}{\ln 2} \sum_{k=1}^{n} W^k[f] k \ln \frac{n}{k} + 3 \cdot \mathrm{Inf}[f]
$$

We will use the following upper bound on W^k:

Theorem 5.2. *[2, 22] Let* $f : \{0, 1\}^n \to \{+1, -1\}$ *be a boolean function. For each* $k \geqslant 2$, *we have*

$$
W^k[f] \leqslant C_k \cdot \mathbb{II}(f) \log^{k-1} \frac{1}{\mathbb{II}(f)}
$$

where C_k *is some constant. Unpublished work of Kindler (cited in [17]) shows that in fact* $C_k \leqslant O(1/k)$.

We will also use the following lower bound on $W^{\leqslant 1}$ of an LTF:

Theorem 5.3. *[7] Let* $f : \{0, 1\}^n \to \{+1, -1\}$ *be a linear threshold function. Then there exist a universal constant* $c > 0$, *such that*

$$
\widehat{f}(\emptyset)^2 + \sum_{i=1}^{n} \widehat{f}(\{i\})^2 \geqslant \frac{1}{2} + c
$$

We call a function f balanced, if $\widehat{f}(\emptyset) = 0$.

Lemma 5.4. *Let* $f:\{0,1\}^n \to \{+1,-1\}$ *be a balanced linear threshold function. Then*

$$\sum_{k=1}^{n} W^k[f] k \ln\frac{n}{k} \leqslant O(\mathbb{III}(f)\cdot \log n).$$

Proof.

$$\begin{aligned}
\sum_{k=1}^{n} W^k[f] k \ln\frac{n}{k} &\leqslant \sum_{k=1}^{n} W^k[f]k\cdot\sum_{j=k}^{n}\frac{1}{j} \leqslant \sum_{j=1}^{n}\frac{1}{j}\sum_{k=1}^{j} W^k[f]k \leqslant c'\cdot\sum_{j=1}^{n}\frac{1}{j}\sum_{k=1}^{j}\mathbb{III}(f)\log^{k-1}\frac{1}{\mathbb{III}(f)}\\
&\leqslant c'\cdot\sum_{j=1}^{n}\frac{1}{j}\mathbb{III}(f)\sum_{k=1}^{\infty}\alpha^{k-1} \quad (\text{using Theorem 5.3, } \alpha := \log\tfrac{1}{\frac{1}{2}+c},\ 0\leqslant\alpha<1.)\\
&\leqslant c'\cdot\mathbb{III}(f)\frac{1}{1-\alpha}\sum_{j=1}^{n}\frac{1}{j}\leqslant c'\cdot\frac{1}{1-\alpha}\cdot\mathbb{III}(f)\cdot(\log n)
\end{aligned}$$

The third inequality follows from Theorem 5.2 and c' is a constant depending on C_k. □

Combining Lemmas 5.1 and 5.4, we obtain the following theorem.

Theorem 5.5. *Let* $f:\{0,1\}^n \to \{+1,-1\}$ *be a balanced linear threshold function. Then,*

$$\mathbb{H}(f) \leqslant \frac{1}{\ln 2}\mathsf{Inf}[f] + 3\cdot\mathsf{Inf}[f] + \frac{c'}{(1-\alpha)\ln 2}\cdot(\log n)\cdot\mathbb{III}(f)$$

Note that since $\mathsf{Inf}_i(f) = |\hat{f}(\{i\})|$ for linear threshold functions $\mathbb{III}(f)$ is at most 1.

5.2 Read-Once Boolean Formulas

It is well-known and easy to see that the mod-2 sum (XOR or Parity) of two functions on disjoint sets of variables simply results in addition of the Fourier entropy and the average sensitivity:

Fact 5.6. *Now, let* $f = g_1 \oplus g_2$ *for* $g_i:\{0,1\}^{V_i}\to\{-1,+1\}$*, where* $V_1\cap V_2=\emptyset$*. Let* $V = V_1\cup V_2$*. Then,*

1. $\mathbb{H}(f) = \mathbb{H}(g_1)+\mathbb{H}(g_2)$
2. $\mathsf{as}(f) = \mathsf{as}(g_1)+\mathsf{as}(g_2)$

We show below somewhat analogous relations when composing functions on disjoint sets of variables using AND and OR operations. The theorem for read-once formulas at once follows from these "tensorizability"-like properties.

For a function $f:\{0,1\}^n\to\{+1,-1\}$ be, we will use $f_{\mathbb{B}}$ to refer to its 0-1 counterpart: $f_{\mathbb{B}} \equiv \frac{1-f}{2}$. In the latter case, we define

$$\mathbf{H}(f_{\mathbb{B}}) = \sum_S \widehat{f_{\mathbb{B}}}^2(S)\log\frac{1}{\widehat{f_{\mathbb{B}}}^2(S)}.$$

Note that this is not exactly an entropy (since we don't normalize), but we will use it simply for notational convenience. An easy relation enables translation between $\mathbb{H}(f)$ and $\mathbf{H}(f_{\mathbb{B}})$:

Lemma 5.7.

$$\mathbb{H}(f) = 4 \cdot \mathbf{H}(f_{\mathbb{B}}) + \varphi(p), \tag{16}$$

where $p = \Pr[f_{\mathbb{B}} = 1] = \widehat{f_{\mathbb{B}}}(\emptyset) = \sum_S \widehat{f_{\mathbb{B}}}^2(S)$, $q := 1 - p$, *and*

$$\varphi(p) := H(4pq) - 4p(H(p) - \log p). \tag{17}$$

Now, let $f = \mathsf{AND}(g_1, g_2)$ for $g_i : \{0,1\}^{V_i} \to \{-1,+1\}$, where $V_1 \cap V_2 = \emptyset$. Let $V = V_1 \cup V_2$. Let $g_{i\mathbb{B}} \equiv \frac{1-g_i}{2}$ and $p_i = \widehat{g_{i\mathbb{B}}}(\emptyset)$. It is then obvious that $f_{\mathbb{B}} \equiv g_{1\mathbb{B}} \cdot g_{2\mathbb{B}}$.

Lemma 5.8. *With the above notations, the following identities hold:*

1. *For all* $S \subseteq V$, $\widehat{f_{\mathbb{B}}}(S) = \widehat{g_{1\mathbb{B}}}(S \cap V_1) \cdot \widehat{g_{2\mathbb{B}}}(S \cap V_2)$
2. $\mathbf{H}(f_{\mathbb{B}}) = p_2 \cdot \mathbf{H}(g_{1\mathbb{B}}) + p_1 \cdot \mathbf{H}(g_{2\mathbb{B}})$
3. $\mathsf{as}(f) = p_2 \cdot \mathsf{as}(g_1) + p_1 \cdot \mathsf{as}(g_2)$

Let us define the following function for $0 \leqslant p \leqslant 1$:

$$\psi(p) := p^2 \log \frac{1}{p^2} - 2\, H(p). \tag{18}$$

The following technical lemma gives us the crucial property of ψ:

Lemma 5.9. *For* ψ *as above and* $p1, p2 \in [0,1]$,

$$p_1 \cdot \psi(p_2) + p_2 \cdot \psi(p_1) \leqslant \psi(p_1 p_2).$$

Proof. We need to prove that $p_1\psi(p_2) + p_2\psi(p_1) - \psi(p_1p_2) \leqslant 0$. Let's begin by manipulating the l.h.s.:

$$\begin{aligned}
&p_1\psi(p_2) + p_2\psi(p_1) - \psi(p_1p_2) \\
&= p_1(p_2^2 \log \frac{1}{p_2^2} - 2\, H(p_2)) + p_2(p_1^2 \log \frac{1}{p_1^2} - 2\, H(p_1)) - (p_1p_2)^2 \log \frac{1}{(p_1p_2)^2} + 2\, H(p_1p_2) \\
&= 2p_1p_2(-p_2 \log p_2 - p_1 \log p_1 + p_1p_2 \log p_2 + p_1p_2 \log p_1) + 2(H(p_1p_2) - p_2\, H(p_1) - p_1\, H(p_2)) \\
&= 2p_1p_2(-p_2q_1 \log p_2 - p_1q_2 \log p_1) + 2(-(1-p_1p_2)\log(1-p_1p_2) + p_2q_1 \log q_1 + p_1q_2 \log q_2) \\
&= 2p_1q_2(-p_1p_2 \log p_1 + \log q_2) + 2p_2q_1(-p_1p_2 \log p_2 + \log q_1) - 2(1-p_1p_2)\log(1-p_1p_2) \\
&\leqslant 2(1-p_1p_2)\left(-p_1p_2 \log(p_1p_2) + \log \frac{q_1q_2}{1-p_1p_2}\right) \quad \text{since } p_1q_2, p_2q_1 \leqslant (1-p_1p_2) \\
&\leqslant 2(1-p_1p_2)\left(-p_1p_2 \log(p_1p_2) + \log \frac{(1-\sqrt{p_1p_2})^2}{(1-p_1p_2)}\right)
\end{aligned}$$

since $q_1q_2 = (1-p_1)(1-p_2) \leqslant (1-\sqrt{p_1p_2})^2$, e.g., by the AM-GM inequality $p_1 + p_2 \geqslant 2\sqrt{p_1p_2}$.

Since $p_1p_2 \in [0,1]$, it suffices to show the (univariate) inequality $\tau(x) := -x\ln x + \ln\frac{(1-\sqrt{x})^2}{1-x} \leqslant 0$ for $x \in [0,1]$. Since the boundary cases are easy to verify, it suffices to prove the that $\tau(x) \leqslant 0$ for $x \in (0,1)$. Note that $\tau(0) = 0$ and hence it suffices to prove that $\tau'(x) < 0$ for $x \in (0,1)$. But

$$\begin{aligned}
\tau'(x) &= -1 + \ln\frac{1}{x} - \frac{1}{\sqrt{x}(1-x)} \\
&\leqslant -1 + \sqrt{\frac{1}{x}} - \frac{1}{\sqrt{x}(1-x)} \quad \text{since } \ln y \leqslant \sqrt{y} \\
&= -1 - \frac{\sqrt{x}}{1-x} \\
&< 0 \quad \text{for } x \in (0,1).
\end{aligned}$$

□

Let us call the following the **FEI01 Inequality:** (Fourier Entropy-Influence Inequality, but the 0-1 version)

$$\mathbf{H}(f_{\mathbb{B}}) \leqslant c \cdot \mathsf{as}(f) + \psi(p), \tag{19}$$

where $p = \widehat{f_{\mathbb{B}}}(\emptyset) = \Pr_x[f_{\mathbb{B}}(x) = 1]$ and c is a constant[3] to be fixed later.

Lemma 5.10. *Suppose $f_{\mathbb{B}} = \mathsf{AND}(g_{1\mathbb{B}}, g_{2\mathbb{B}})$, where the g_i depend on* disjoint *sets of variables. If each of the g_i satisfies the FEI01 Inequality (19), then so does f.*

Proof.

$$\begin{aligned}
\mathbf{H}(f_{\mathbb{B}}) &= p_2\,\mathbf{H}(g_{1\mathbb{B}}) + p_1\,\mathbf{H}(g_{2\mathbb{B}}) && \text{by Lemma 5.8(2)} \\
&\leqslant p_2(c\cdot \mathsf{as}(g_1) + \psi(p_1)) + p_1(c\cdot\mathsf{as}(g_2) + \psi(p_2)) && \text{since } g_i \text{ satisfy (19)} \\
&= c\cdot(p_2\mathsf{as}(g_1) + p_1\mathsf{as}(g_2)) + (p_2\psi(p_1) + p_1\psi(p_2)) \\
&\leqslant c\cdot\mathsf{as}(f) + \psi(p) && \text{by Lemma 5.8(3) and Lemma 5.9}
\end{aligned}$$

□

Lemma 5.11. *If f satisfies FEI01 inequality (19), then so does its negation, i.e., $1-f$.*

Proof. Note that $\mathbf{H}(1-f) = \mathbf{H}(f) - p^2\log\frac{1}{p^2} + q^2\log\frac{1}{q^2}$ and $\psi(p) - \psi(q) = p^2\log\frac{1}{p^2} - q^2\log\frac{1}{q^2}$ (because $H(p) = H(q)$). □

Corollary 5.12. *Suppose $f_{\mathbb{B}} = \mathsf{OR}(g_{1\mathbb{B}}, g_{2\mathbb{B}})$, where the g_i depend on* disjoint *sets of variables. If each of the g_i satisfies the FEI01 Inequality (19), then so does f.*

Proof. Note that $1 - f_{\mathbb{B}} = (1-g_{1\mathbb{B}})\cdot(1-g_{2\mathbb{B}})$ and apply lemmas 5.10 and 5.11. □

Theorem 5.13. *The FEI01 inequality holds for all read-once Boolean formulas. In fact, inequality (19) is satisfied for any read-once formula f with constant $c = 5/2$.*

[3] We do not attempt to find optimal constants here. By fine-tuning the choices of various functions, it may be possible to get better constants.

Proof. Let f be computed by a read-once Boolean formula. We proceed by induction on the tree. At the leaves f is a single variable, say x_1. Then $\mathbf{H}(f_{\mathbb{B}}) = \frac{1}{4}\log 4 + \frac{1}{4}\log 4 = \frac{1}{2}$ since $f_{\mathbb{B}}(\emptyset) = 1/2$ and $f_{\mathbb{B}}(\{1\}) = -1/2$, $\mathsf{as}(f) = 1$, $p = 1/2$, and $\psi(1/2) = -3/2$. Thus with $c = 5/2$, (19) is satisfied.

Now, lemma 5.10 and 5.12 imply that at every AND gate and OR gate, the inequality (19) is preserved, i.e., if it holds at both the inputs, it also holds at the output. □

Combining this theorem with (16) and (19) and noting that $4\psi(p) + \varphi(p) \leqslant 0$ for $p \in [0, 1]$, we conclude that the Fourier Entropy Influence conjecture (2) holds for read-once formulas:

Corollary 5.14. *If f is computed by a read-once formula, then* $\mathbb{H}(f) \leqslant 10\,\mathsf{Inf}(f)$.

5.2.1 Extension to Read-Once Formulas with Parity Gates

We show below that the above result can be extended to show the Entropy Influence Conjecture holds for read-once formulas that include parity gates at internal nodes (in addition to AND and OR). Along the same lines as before, we formulate an inequality for the Fourier entropy that is preserved by the XOR operator using Fact 5.6. Specifically, consider the inequality

$$\mathbb{H}(f) \leqslant 10 \cdot \mathsf{as}(f) + \kappa(p), \tag{20}$$

where

$$\kappa(p) := 4\psi(p) + \varphi(p) = -8\,H(p) - 8pq - (1 - 4pq)\log(1 - 4pq). \tag{21}$$

We already know that (20) is preserved by AND and OR operators. It suffices, therefore, to show that it is also preserved by the parity operator. This will follow if we can show that

$$\kappa(p_1) + \kappa(p_2) \leqslant \kappa(p_1q_2 + p_2q_1), \tag{22}$$

since if $f = g_1 \cdot g_2$ (note that we are back to $\{-1, +1\}$-valued Boolean functions and hence parity is simply product) then $p = p_1q_2 + p_2q_1$. Lemma 5.17 below proves this property of κ. Assuming this we prove the following lemma which leads us to the main theorem of this section.

Lemma 5.15. *Suppose $f = g_1 \cdot g_2$, where the g_i depend on* disjoint *sets of variables. If each of the g_i satisfies the entropy-influence inequality (20), then so does f.*

Proof.

$$\begin{aligned}
\mathbb{H}(f) &= \mathbb{H}(g_1) + \mathbb{H}(g_2) && \text{by Fact 5.6(i)}\\
&\leqslant 10 \cdot \mathsf{as}(g_1) + \kappa(p_1) + 10 \cdot \mathsf{as}(g_2) + \kappa(g_2) && \text{since } g_i \text{ satisfy (20)}\\
&= 10 \cdot \mathsf{as}(f) + \kappa(p_1) + \kappa(p_2) && \text{by Fact 5.6(ii)}\\
&\leqslant 10 \cdot \mathsf{as}(f) + \kappa(p) && \text{by Lemma 5.17 below}
\end{aligned}$$

□

Theorem 5.16. *If* f *is computed by a read-once formula using* AND*,* OR*, and* XOR *gates, then* $\mathbb{H}(f) \leqslant 10 \operatorname{Inf}(f)$*.*

Proof. We use induction on the tree given by the formula computing f to prove (20). The leaves are input variables or their negations and the claim that they satisfy (20) can be verified by direct calculation. At any internal node, its two inputs are given by subformulas depending on disjoint sets of variables by the read-once property of the formula. When the internal node is an AND or OR gate, the claim follows from Lemma 5.10, Corollary 5.12, (16), and (21). When the internal node is an XOR gate, the claim follows from Lemma 5.15. Thus (20) holds at the root of the tree and hence for f. It is easy to verify that $-10 \leqslant \kappa(p) \leqslant 0$ for $p \in [0, 1]$. This proves the theorem. □

To complete the above proof, we still need to prove the property (22) of κ. We do this below.

Lemma 5.17. *For* κ *as defined by* (21)*,* $\kappa(p_1) + \kappa(p_2) \leqslant \kappa(p_1q_2 + p_2q_1)$*.*

Proof. In the following, we will let $p = p_1q_2 + p_2q_1$ and $q = 1 - p = p_1p_2 + q_1q_2$.

To begin with, we observe that $(1-4pq) = (p-q)^2$ and that $(p-q) = (p_1 - q_1)(p_2 - q_2)$, i.e., parity operation on independent Boolean variables results in multiplying their *biases*, and hence $(1 - 4pq) = (1 - 4p_1q_1)(1 - 4p_2q_2)$. Using this, we relate the third terms on either side of the inequality to be proved.

$$\begin{aligned}
(1-4pq)\log(1-4pq) &= (1-4p_1q_1)(1-4p_2q_2)\log((1-4p_1q_1)1-4p_2q_2)) \\
&= (1-4p_2q_2)\left((1-4p_1q_1)\log(1-4p_1q_1)\right) + (1-4p_1q_1)\left((1-4p_2q_2)\log(1-4p_2q_2)\right) \\
&\leqslant \left((1-4p_1q_1)\log(1-4p_1q_1)\right) + \left((1-4p_2q_2)\log(1-4p_2q_2)\right) + 64p_1q_1p_2q_2,
\end{aligned}$$

using that $-(1-4p_1q_1)\log(1-4p_1q_1) \leqslant 8p_1q_1$ (this follows from the inequality $x \log \frac{1}{x} \leqslant$ [4] $2(1-x)$ for $x \in [0, 1]$). Thus, we have

$$-(1-4p_1q_1)\log(1-4p_1q_1) - (1-4p_2q_2)\log(1-4p_2q_2) + (1-4pq)\log(1-4pq) \leqslant 64p_1q_1p_2q_2. \tag{23}$$

Next, we simplify the second terms:

$$\begin{aligned}
pq = (p_1q_2 + p_2q_1)(p_1p_2 + q_1q_2) &= p_1q_1(p_2^2 + q_2^2) + p_2q_2(p_1^2 + q_1^2) \\
&= p_1q_1(1 - 2p_2q_2) + p_2q_2(1 - 2p_1q_1) \\
&= p_1q_1 + p_2q_2 - 4p_1q_1p_2q_2.
\end{aligned}$$

Hence, we have

$$-8p_1q_1 - 8p_2q_2 + 8pq = -32p_1q_1p_2q_2. \tag{24}$$

[4] Any constant $c \geqslant \frac{1}{\ln 2}$ can be used instead of 2

Finally, the first terms:

$$\begin{aligned}
H(p) &= H(p_1q_2 + p_2q_1) \\
&= (p_1q_2 + p_2q_1)\log\frac{1}{(p_1q_2+p_2q_1)} + (p_1p_2+q_1q_2)\log\frac{1}{(p_1p_2+q_1q_2)} \\
&= p_1q_2\log\frac{1}{p_1q_2} + p_1q_2\log\frac{p_1q_2}{(p_1q_2+p_2q_1)} + p_2q_1\log\frac{1}{p_2q_1} + p_2q_1\log\frac{p_2q_1}{(p_1q_2+p_2q_1)} \\
&\quad + \text{ similar terms for the second summand} \\
&= q_2(-p_1\log p_1) + p_1(-q_2\log q_2) + p_2(-q_1\log q_1) + q_1(-p_2\log p_2) \\
&\quad + p_1q_2\log\frac{p_1q_2}{(p_1q_2+p_2q_1)} + p_2q_1\log\frac{p_2q_1}{(p_1q_2+p_2q_1)} \\
&\quad + \text{ similar terms from the second half} \\
&= -p_1\log p_1(q_2+p_2) - q_1\log q_1(p_2+q_2) - p_2\log p_2(q_1+p_1) - q_2\log q_2(p_1+q_1) \\
&\quad + p_1q_2\log\frac{p_1q_2}{(p_1q_2+p_2q_1)} + p_2q_1\log\frac{p_2q_1}{(p_1q_2+p_2q_1)} + p_1p_2\log\frac{p_1p_2}{(p_1p_2+q_1q_2)} + q_1q_2\log\frac{q_1q_2}{(p_1p_2+q_1q_2)} \\
&= H(p_1) + H(p_2) - (p_1q_2+p_2q_1)\,H\left(\frac{p_1q_2}{(p_1q_2+p_2q_1)}\right) - (p_1p_2+q_1q_2)\,H\left(\frac{p_1p_2}{(p_1p_2+q_1q_2)}\right) \\
&\leqslant H(p_1) + H(p_2) - 2\min\{p_1q_2, p_2q_1\} - 2\min\{p_1p_2, q_1, q_2\} \quad \text{using } H(p) \geqslant 2\min\{p,q\} \\
&\leqslant H(p_1) + H(p_2) - 2p_1q_2p_2q_1 - 2p_1p_2q_1q_2 \quad \text{since } \min\{p,q\} \geqslant pq \text{ for } 0 \leqslant p,q \leqslant 1 \\
&= H(p_1) + H(p_2) - 4p_1q_1p_2q_2.
\end{aligned}$$

Hence, we have

$$-8\,H(p_1) - 8\,H(p_2) + 8\,H(p) \leqslant -32p_1q_1p_2q_2. \tag{25}$$

Combing (23), (24), (25), and the definition (21), we obtain

$$\kappa(p_1) + \kappa(p_2) - \kappa(p) \leqslant 0$$

and this concludes the proof. □

References

[1] Michael Ben-Or and Nathan Linial. Collective coin flipping. In Silvio Micali, editor, *Randomness and Computation*. Academic Press, New York, 1990.

[2] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise sensitivity of boolean functions and applications to percolation. *Publications Mathmatiques de l'Institut des Hautes tudes Scientifiques*, 90(1):5–43, 1999.

[3] Ravi Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997.

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

[5] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288(1):21–43, 2002.

[6] Bireswar Das, Manjish Pal, and Vijay Visavaliya. The entropy influence conjecture revisited. Technical report, arXiv:1110.4301, 2011.

[7] Anindya De, Ilias Diakonikolas, and Rocco A. Servedio. Robust khintchine-kahane inequality and computing optimal constants in fourier analysis and high-dimensional geometry. *CoRR*, abs/1207.2229, 2012.

[8] Ronald de Wolf. A brief introduction to fourier analysis on the boolean cube. *Theory of Computing, Graduate Surveys*, 1:1–20, 2008.

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

[10] 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*, STOC '08, pages 527–536, 2008.

[11] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[12] Nathan Keller, Elchanan Mossel, and Tomer Schlank. A note on the entropy/influence conjecture. Technical report, arXiv:1105.2651, 2011.

[13] Adam Klivans, Homin Lee, and Andrew Wan. Mansour's conjecture is true for random dnf formulas. In *Proceedings of the 23rd Conference on Learning Theory*, pages 368–380, 2010.

[14] Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the fourier spectrum. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, pages 455–464, 1991.

[15] Yishay Mansour. Learning boolean functions via the fourier transform. pages 391–424, 1994.

[16] Yishay Mansour. An $o(n^{\log\log n})$ learning algorithm for dnf under the uniform distribution. *Journal of Computer and System Sciences*, 50(3):543–550, 1995.

[17] Ryan O'Donnell. The lecture notes of the course "analysis of boolean functions": Lecture 29: Open problems, 2007.

[18] Ryan O'Donnell. Some topics in analysis of boolean functions. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 569–578, 2008.

[19] Ryan O'Donnell. Analysis of boolean functions, 2012.

[20] Ryan O'Donnell, John Wright, and Yuan Zhou. The fourier entropy-influence conjecture for certain classes of boolean functions. In *Proceedings of Automata, Languages and Programming - 38th International Colloquium*, pages 330–341, 2011.

[21] Michel Talagrand. On russo's approximate zero-one law. *Annals of Probability*, 22(3):1576–1587, 1994.

[22] Michel Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de