A composition theorem for the Fourier Entropy-Influence conjecture

Ryan O'Donnell[1] ⋆ and Li-Yang Tan[2] ⋆⋆

[1] Carnegie Mellon University
[2] Columbia University

Abstract. The Fourier Entropy-Influence (FEI) conjecture of Friedgut and Kalai [1] seeks to relate two fundamental measures of Boolean function complexity: it states that $\mathbf{H}[f] \leq C \cdot \mathbf{Inf}[f]$ holds for every Boolean function f, where $\mathbf{H}[f]$ denotes the spectral entropy of f, $\mathbf{Inf}[f]$ is its total influence, and $C > 0$ is a universal constant. Despite significant interest in the conjecture it has only been shown to hold for a few classes of Boolean functions.

Our main result is a composition theorem for the FEI conjecture. We show that if $g_1, \ldots, g_k$ are functions over disjoint sets of variables satisfying the conjecture, and if the Fourier transform of F taken with respect to the product distribution with biases $\mathbf{E}[g_1], \ldots, \mathbf{E}[g_k]$ satisfies the conjecture, then their composition $F(g_1(x^1), \ldots, g_k(x^k))$ satisfies the conjecture. As an application we show that the FEI conjecture holds for read-once formulas over arbitrary gates of bounded arity, extending a recent result [2] which proved it for read-once decision trees. Our techniques also yield an explicit function with the largest known ratio of $C \geq 6.278$ between $\mathbf{H}[f]$ and $\mathbf{Inf}[f]$, improving on the previous lower bound of 4.615.

1 Introduction

A longstanding and important open problem in the field of Analysis of Boolean Functions is the Fourier Entropy-Influence conjecture made by Ehud Friedgut and Gil Kalai in 1996 [1,3]. The conjecture seeks to relate two fundamental analytic measures of Boolean function complexity, the spectral entropy and total influence:

Fourier Entropy-Influence (FEI) Conjecture. *There exists a universal constant $C > 0$ such that for every Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$, it holds*

⋆ Supported by NSF grants CCF-0747250 and CCF-1116594, and a Sloan fellowship. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

⋆⋆ Research done while visiting CMU.

that $\mathbf{H}[f] \le C \cdot \mathbf{Inf}[f]$. *That is,*

$$\sum_{S \subseteq [n]} \widehat{f}(S)^2 \log_2\left(\frac{1}{\widehat{f}(S)^2}\right) \le C \sum_{S \subseteq [n]} |S| \cdot \widehat{f}(S)^2.$$

Applying Parseval's identity to a Boolean function f we get $\sum_{S\subseteq[n]} \widehat{f}(S)^2 = \mathbf{E}[f(\boldsymbol{x})^2] = 1$, and so the Fourier coefficients of f induce a probability distribution $\mathcal{S}_f$ over the 2^n subsets of $[n]$ wherein $S \subseteq [n]$ has "weight" (probability mass) $\widehat{f}(S)^2$. The *spectral entropy* of f, denoted $\mathbf{H}[f]$, is the Shannon entropy of $\mathcal{S}_f$, quantifying how spread out the Fourier weight of f is across all 2^n monomials. The influence of a coordinate $i \in [n]$ on f is $\mathbf{Inf}_i[f] = \mathbf{Pr}[f(\boldsymbol{x}) \neq f(\boldsymbol{x}^{\oplus i})]$[3], where $\boldsymbol{x}^{\oplus i}$ denotes $\boldsymbol{x}$ with its i-th bit flipped, and the *total influence* of f is simply $\mathbf{Inf}[f] = \sum_{i=1}^n \mathbf{Inf}_i[f]$. Straightforward Fourier-analytic calculations show that this combinatorial definition is equivalent to the quantity $\mathbf{E}_{\mathbf{S}\sim\mathcal{S}_f}[|\mathbf{S}|] = \sum_{S\subseteq[n]} |S| \cdot \widehat{f}(S)^2$, and so total influence measures the degree distribution of the monomials of f, weighted by the squared-magnitude of its coefficients. Roughly speaking then, the FEI conjecture states that a Boolean function whose Fourier weight is well "spread out" (*i.e.* has high spectral entropy) must have a significant portion of its Fourier weight lying on high degree monomials (*i.e.* have high total influence).[4]

In addition to being a natural question concerning the Fourier spectrum of Boolean functions, the FEI conjecture also has important connections to several areas of theoretical computer science and mathematics. Friedgut and Kalai's original motivation was to understand general conditions under which monotone graph properties exhibit sharp thresholds, and the FEI conjecture captures the intuition that having significant symmetry, hence high spectral entropy, is one such condition. Besides its applications in the study of random graphs, the FEI conjecture is known to imply the celebrated Kahn-Kalai-Linial theorem [4]:

KKL Theorem. *For every Boolean function* f *there exists an* $i \in [n]$ *such that* $\mathbf{Inf}_i[f] = \mathbf{Var}[f] \cdot \Omega(\frac{\log n}{n})$.

The FEI conjecture also implies Mansour's conjecture [5]:

Mansour's Conjecture. *Let* f *be a Boolean function computed by a* t*-term DNF formula. For any constant* $\varepsilon > 0$ *there exists a collection* $\mathcal{S} \subseteq 2^{[n]}$ *of cardinality* $\mathrm{poly}(t)$ *such that* $\sum_{S\in\mathcal{S}} \widehat{f}(S)^2 \ge 1 - \varepsilon$.

Combined with recent work of Gopalan *et al.* [6], Mansour's conjecture yields an efficient algorithm for agnostically learning the class of $\mathrm{poly}(n)$-term DNF

[3] All probabilities and expectations are with respect to the uniform distribution unless otherwise stated.

[4] The assumption that f is Boolean-valued is crucial here, as the same conjecture is false for functions $f : \{-1,1\}^n \to \mathbb{R}$ satisfying $\sum_{S\subseteq[n]} \widehat{f}(S)^2 = 1$. The canonical counterexample is $f(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i$ which has total influence 1 and spectral entropy $\log_2 n$.

formulas from queries. This would resolve a central open problem in computational learning theory [7]. De *et al.* also noted that sufficiently strong versions of Mansour's conjecture would yield improved pseudorandom generators for depth-2 AC^0 circuits [8]. More generally, the FEI conjecture implies the existence of sparse L_2-approximators for Boolean functions with small total influence:

Sparse L_2-approximators. Assume the FEI conjecture holds. Then for every Boolean function f there exists a $2^{O(\mathbf{Inf}[f]/\varepsilon)}$-sparse polynomial $p : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbf{E}[(f(\boldsymbol{x}) - p(\boldsymbol{x}))^2] \leq \varepsilon$.

By Friedgut's junta theorem [9], the above holds unconditionally with a weaker bound of $2^{O(\mathbf{Inf}[f]^2/\varepsilon^2)}$. This is the main technical ingredient underlying several of the best known uniform-distribution learning algorithms [10,11].

For more on the FEI conjecture we refer the reader to Kalai's blog post [3].

1.1 Our results

Our research is motivated by the following question:

Question 1. Let $F : \{-1,1\}^k \to \{-1,1\}$ and $g_1, \ldots, g_k : \{-1,1\}^\ell \to \{-1,1\}$. What properties do F and $g_1, \ldots, g_k$ have to satisfy for the FEI conjecture to hold for the disjoint composition $f(x^1, \ldots, x^k) = F(g_1(x^1), \ldots, g_k(x^k))$?

Despite its simplicity this question has not been well understood. For example, prior to our work the FEI conjecture was open even for read-once DNFs (such as the "tribes" function); these are the disjoint compositions of $F = \mathsf{OR}$ and $g_1, \ldots, g_k = \mathsf{AND}$, perhaps two of the most basic Boolean functions with extremely simple Fourier spectra. Indeed, Mansour's conjecture, a weaker conjecture than FEI, was only recently shown to hold for read-once DNFs [12,8]. Besides being a fundamental question concerning the behavior of spectral entropy and total influence under composition, Question 1 (and our answer to it) also has implications for a natural approach towards disproving the FEI conjecture; we elaborate on this at the end of this section.

A particularly appealing and general answer to Question 1 that one may hope for would be the following: "if $\mathbf{H}[F] \leq C_1 \cdot \mathbf{Inf}[F]$ and $\mathbf{H}[g_i] \leq C_2 \cdot \mathbf{Inf}[g_i]$ for all $i \in [k]$, then $\mathbf{H}[f] \leq \max\{C_1, C_2\} \cdot \mathbf{Inf}[f]$." While this is easily seen to be false[5], our main result shows that this proposed answer to Question 1 *is* in fact true for a carefully chosen sharpening of the FEI conjecture. To arrive at a formulation that bootstraps itself, we first consider a slight strengthening of the FEI conjecture which we call FEI$^+$, and then work with a generalization of FEI$^+$ that concerns the Fourier spectrum of f not just with respect to the uniform distribution, but an arbitrary product distribution over $\{-1,1\}^n$:

[5] For example, by considering $F = \mathsf{OR}_2$, the 2-bit disjunction, and $g_1, g_2 = \mathsf{AND}_2$, the 2-bit conjunction.

Conjecture 1 (FEI$^+$ for product distributions). There is a universal constant $C > 0$ such that the following holds. Let $\mu = \langle \mu_1, \dots, \mu_n \rangle$ be any sequence of biases and $f : \{-1,1\}^n_\mu \to \{-1,1\}$. Here the notation $\{-1,1\}^n_\mu$ means that we think of $\{-1,1\}^n$ as being endowed with the μ-biased product probability distribution in which $\mathbf{E}_\mu[x_i] = \mu_i$ for all $i \in [n]$. Let $\{\widetilde{f}(S)\}_{S \subseteq [n]}$ be the μ-biased Fourier coefficients of f. Then

$$\sum_{S \neq \emptyset} \widetilde{f}(S)^2 \log \left(\frac{\prod_{i \in S}(1-\mu_i^2)}{\widetilde{f}(S)^2} \right) \leq C \cdot (\mathbf{Inf}^\mu[f] - \underset{\mu}{\mathbf{Var}}[f]).$$

We write $\mathbf{H}^\mu[f]$ to denote the quantity $\sum_{S \subseteq [n]} \widetilde{f}(S)^2 \log \left(\prod_{i \in S}(1-\mu_i^2)/\widetilde{f}(S)^2 \right)$, and so the inequality of Conjecture 1 can be equivalently stated as $\mathbf{H}^\mu[f^{\geq 1}] \leq C \cdot (\mathbf{Inf}^\mu[f] - \mathbf{Var}_\mu[f])$.

In Proposition 1 we show that Conjecture 1 with $\mu = \langle 0, \dots, 0 \rangle$ (the uniform distribution) implies the FEI conjecture. We say that a Boolean function f "satisfies μ-biased FEI$^+$ with factor C" if the μ-biased Fourier transform of f satisfies the inequality of Conjecture 1. Our main result, which we prove in Section 3, is a composition theorem for FEI$^+$:

Theorem 1. *Let $f(x^1, \dots, x^k) = F(g_1(x^1), \dots, g_k(x^k))$, where the domain of f is endowed with a product distribution μ. Suppose $g_1, \dots, g_k$ satisfy μ-biased FEI$^+$ with factor C_1 and F satisfies η-biased FEI$^+$ with factor C_2, where $\eta = \langle \mathbf{E}_\mu[g_1], \dots, \mathbf{E}_\mu[g_k] \rangle$. Then f satisfies μ-biased FEI$^+$ with factor* $\max\{C_1, C_2\}$.

Theorem 1 suggests an inductive approach towards proving the FEI conjecture for read-once de Morgan formulas: since the dictators $\pm x_i$ trivially satisfy uniform-distribution FEI$^+$ with factor 1, it suffices to prove that both AND_2 and OR_2 satisfy μ-biased FEI$^+$ with some constant *independent of* $\mu \in [-1,1]^2$. In Section 4 we prove that in fact *every* $F : \{-1,1\}^k \to \{-1,1\}$ satisfies μ-biased FEI$^+$ with a factor depending only on its arity k and not the biases $\mu_1, \dots, \mu_k$.

Theorem 2. *Every $F : \{-1,1\}^k \to \{-1,1\}$ satisfies μ-biased FEI$^+$ with factor $C = 2^{O(k)}$ for any product distribution $\mu = \langle \mu_1, \dots, \mu_k \rangle$.*

Together, Theorems 1 and 2 imply:

Theorem 3. *Let f be computed by a read-once formula over the basis $\mathcal{B}$ and μ be any sequences of biases. Then f satisfies μ-biased FEI$^+$ with factor C, where C depends only on the arity of the gates in $\mathcal{B}$.*

Since uniform-distribution FEI$^+$ is a strengthening of the FEI conjecture, Theorem 3 implies that the FEI conjecture holds for read-once formulas over arbitrary gates of bounded arity. As mentioned above, prior to our work the FEI conjecture was open even for the class of read-once DNFs, a small subclass of read-once formulas over the de Morgan basis $\{\mathsf{AND}_2, \mathsf{OR}_2, \mathsf{NOT}\}$ of arity 2. Read-once formulas over a rich basis $\mathcal{B}$ are a natural generalization of read-once de Morgan formulas, and have seen previous study in concrete complexity (see *e.g.* [13]).

Improved lower bound on the FEI constant. Iterated disjoint composition is commonly used to achieve separations between complexity measures for Boolean functions [14], and represents a natural approach towards disproving the FEI conjecture. For example, one may seek a function F such that iterated compositions of F with itself achieves a super-constant amplification of the ratio between $\mathbf{H}[F]$ and $\mathbf{Inf}[F]$, or consider variants such as iterating F with a different combining function G. Theorem 3 rules out as potential counterexamples all such constructions based on iterated composition.

However, the tools we develop to prove Theorem 3 also yield an explicit function f achieving the best-known separation between $\mathbf{H}[f]$ and $\mathbf{Inf}[f]$ (*i.e.* the constant C in the statement of the FEI conjecture). In Section 5 we prove:

Theorem 4. *There exists an explicit family of functions $f_n : \{-1,1\}^n \to \{-1,1\}$ such that*

$$\lim_{n\to\infty} \frac{\mathbf{H}[f_n]}{\mathbf{Inf}[f_n]} \geq 6.278.$$

This improves on the previous lower bound of $C \geq 60/13 \approx 4.615$ [2].

Previous work. The first published progress on the FEI conjecture was by Klivans *et al.* who proved the conjecture for random poly(n)-term DNF formulas [12]. This was followed by the work of O'Donnell *et al.* who proved the conjecture for the class of symmetric functions and read-once decision trees [2].

The FEI conjecture for product distributions was studied in the recent work of Keller *et al.* [15], where they consider the case of all the biases being the same. They introduce the following generalization of the FEI conjecture to these measures, and show via a reduction to the uniform distribution [16] that it is equivalent to the FEI conjecture:

Conjecture 2 (Keller-Mossel-Schlank). There is a universal constant C such that the following holds. Let $0 < p < 1$ and $f : \{-1,1\}^n \to \{-1,1\}$, where the domain of f is endowed with the product distribution where $\mathbf{Pr}[x_i = -1] = p$ for all $i \in [n]$. Let $\{\widetilde{f}(S)\}_{S\subseteq[n]}$ be the Fourier coefficients of f with respect to this distribution. Then

$$\sum_{S\subseteq[n]} \widetilde{f}(S)^2 \log_2\left(\frac{1}{\widetilde{f}(S)^2}\right) \leq C \cdot \frac{\log(1/p)}{1-p} \sum_{S\subseteq[n]} |S| \cdot \widetilde{f}(S)^2.$$

Notice that in this conjecture, the constant on the right-hand side, $C\cdot\frac{\log(1/p)}{1-p}$, depends on p. By way of contrast, in our Conjecture 1 the right-hand side constant has no dependence on p; instead, the dependence on the biases is built into the definition of spectral entropy. We view our generalization of the FEI conjecture to arbitrary product distributions (where the biases are not necessarily identical) as a key contribution of this work, and point to our composition theorem as evidence in favor of Conjecture 1 being a good statement to work with.

2 Preliminaries

Notation. We will be concerned with functions $f : \{-1,1\}^n_\mu \to \mathbb{R}$ where $\mu = \langle \mu_1, \ldots, \mu_n \rangle \in [0,1]^n$ is a sequence of biases. Here the notation $\{-1,1\}^n_\mu$ means that we think of $\{-1,1\}^n$ as being endowed with the μ-biased product probability distribution in which $\mathbf{E}_\mu[x_i] = \mu_i$ for all $i \in [n]$. We write σ_i^2 to denote variance of the i-th coordinate $\mathbf{Var}_\mu[x_i] = 1 - \mu_i^2$, and $\varphi : \mathbb{R} \to \mathbb{R}$ as shorthand for the function $t \mapsto t^2 \log(1/t^2)$, adopting the convention that $\varphi(0) = 0$. We will assume familiarity with the basics of Fourier analysis with respect to product distributions over $\{-1,1\}^n$; a review is included in Appendix A.

Proposition 1 (FEI$^+$ implies FEI). *Suppose f satisfies uniform-distribution FEI$^+$ with factor C. Then f satisfies the FEI conjecture with factor* $\max\{C, 1/\ln 2\}$.

Proof. Let $\widehat{f}(\emptyset)^2 = 1 - \varepsilon$, where $\varepsilon = \mathbf{Var}[f]$ by Parseval's identity. By our assumption that f satisfies uniform-distribution FEI$^+$ with factor C, we have

$$\begin{aligned}\sum_{S \subseteq [n]} \widehat{f}(S)^2 \log\left(\frac{\prod_{i\in S}\sigma_i^2}{\widehat{f}(S)^2}\right) &\le C \cdot (\mathbf{Inf}[f] - \mathbf{Var}[f]) + (1-\varepsilon)\log\frac{1}{(1-\varepsilon)} \\ &\le C \cdot (\mathbf{Inf}[f] - \mathbf{Var}[f]) + \frac{\varepsilon}{\ln 2} \\ &= C \cdot \mathbf{Inf}[f] + \left(\frac{1}{\ln 2} - C\right) \cdot \mathbf{Var}[f].\end{aligned}$$

If $C > 1/\ln 2$ then the RHS is at most $C \cdot \mathbf{Inf}[f]$ since $(\frac{1}{\ln 2} - C) \cdot \mathbf{Var}[f]$ is negative. Otherwise we apply the Poincaré inequality (Theorem 7) to conclude that the RHS is at most $C \cdot \mathbf{Inf}[f] + (\frac{1}{\ln 2} - C) \cdot \mathbf{Inf}[f] = \frac{1}{\ln 2} \cdot \mathbf{Inf}[f]$. $\square$

3 Composition theorem for FEI$^+$

We will be concerned with compositions of functions $f = F(g_1(x^1), \ldots, g_k(x^k))$ where $g_1, \ldots, g_k$ are over disjoint sets of variables each of size ℓ. The domain of each g_i is endowed with a product distribution $\mu^i = \langle \mu_1^i, \ldots, \mu_\ell^i \rangle$, which induces an overall product distribution $\mu = \langle \mu_1^1, \ldots, \mu_\ell^1, \ldots, \mu_1^k, \ldots, \mu_\ell^k \rangle$ over the domain of $f : \{-1,1\}^{k\ell} \to \{-1,1\}$. For notational clarity we will adopt the equivalent view of $g_1, \ldots, g_k$ as functions over the same domain $\{-1,1\}^{k\ell}_\mu$ endowed with the same product distribution μ, with each g_i depending only on ℓ out of $k\ell$ variables.

Our first lemma gives formulas for the spectral entropy and total influence of the product of functions $\Phi_1, \ldots, \Phi_k$ over disjoint sets of variables. The lemma holds for real-valued functions Φ_i; we require this level of generality as we will not be applying the lemma directly to the Boolean-valued functions $g_1, \ldots, g_k$ in the composition $F(g_1(x^1), \ldots, g_k(x^k))$, but instead to their normalized variants $\Phi(g_i) = (g_i - \mathbf{E}[g_i])/\,\mathbf{Var}[g_i]^{1/2}$.

Lemma 1. *Let $\Phi_1, \ldots, \Phi_k : \{-1,1\}^{k\ell}_\mu \to \mathbb{R}$ where each Φ_i depends only on the ℓ coordinates in $\{(i-1)\ell+1, \ldots, i\ell\}$. Then*

$$\mathbf{H}^\mu[\Phi_1 \cdots \Phi_k] = \sum_{i=1}^k \mathbf{H}^\mu[\Phi_i] \prod_{j \neq i} \mathop{\mathbf{E}}_\mu[\Phi_j^2] \text{ and } \mathbf{Inf}^\mu[\Phi_1 \cdots \Phi_k] = \sum_{i=1}^k \mathbf{Inf}^\mu[\Phi_i] \prod_{j \neq i} \mathop{\mathbf{E}}_\mu[\Phi_j^2].$$

Due to space considerations we defer the proof of Lemma 1 to Appendix B. We note that this lemma recovers as a special case the folklore observation that the FEI conjecture "tensorizes": for any f if we define $f^{\oplus k}(x^1, \ldots, x^k) = f(x^1) \cdots f(x^k)$ then $\mathbf{H}[f^{\oplus k}] = k \cdot \mathbf{H}[f]$ and $\mathbf{Inf}[f^{\oplus k}] = k \cdot \mathbf{Inf}[f]$. Therefore $\mathbf{H}[f] \leq C \cdot \mathbf{Inf}[f]$ if and only if $\mathbf{H}[f^{\oplus k}] \leq C \cdot \mathbf{Inf}[f^{\oplus k}]$.

Our next proposition relates the basic analytic measures – spectral entropy, total influence, and variance – of a composition $f = F(g_1(x^1), \ldots, g_k(x^k))$ to the corresponding quantities of the combining function F and base functions $g_1, \ldots, g_k$. As alluded to above, we accomplish this by considering f as a linear combination of the normalized functions $\Phi(g_i) = (g_i - \mathbf{E}[g_i])/\,\mathbf{Var}[g_i]^{1/2}$ and applying Lemma 1 to each term in the sum. We mention that this proposition is also the crux of our new lower bound of $C \geq 6.278$ on the constant of the FEI conjecture, which we present in Section 5.

Proposition 2. *Let $F : \{-1,1\}^k \to \mathbb{R}$, and $g_1, \ldots, g_k : \{-1,1\}^{k\ell}_\mu \to \{-1,1\}$ where each g_i depends only on the ℓ coordinates in $\{(i-1)\ell+1, \ldots, i\ell\}$. Let $f(x) = F(g_1(x), \ldots, g_k(x))$ and $\{\widetilde{F}(S)\}_{S \subseteq [k]}$ be the η-biased Fourier coefficients of F where $\eta = \langle \mathbf{E}_\mu[g_1]), \ldots, \mathbf{E}_\mu[g_k] \rangle$. Then*

$$\mathbf{H}^\mu[f^{\geq 1}] = \mathbf{H}^\eta[F^{\geq 1}] + \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \frac{\mathbf{H}^\mu[g_i^{\geq 1}]}{\mathbf{Var}_\mu[g_i]}, \tag{1}$$

$$\mathbf{Inf}^\mu[f] = \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \frac{\mathbf{Inf}^\mu[g_i]}{\mathbf{Var}_\mu[g_i]}, \quad \textit{and} \tag{2}$$

$$\mathbf{Var}_\mu[f] = \sum_{S \neq \emptyset} \widetilde{F}(S)^2 = \mathbf{Var}_\eta[F]. \tag{3}$$

Proof. By the η-biased Fourier expansion of $F : \{-1,1\}^k_\eta \to \mathbb{R}$ and the definition of η we have

$$F(y_1, \ldots, y_k) = \sum_{S \subseteq [n]} \widetilde{F}(S) \prod_{i \in S} \frac{y_i - \eta_i}{\sqrt{1 - \eta_i^2}} = \sum_{S \subseteq [n]} \widetilde{F}(S) \prod_{i \in S} \frac{y_i - \mathbf{E}_\mu[g_i]}{\mathbf{Var}_\mu[g_i]^{1/2}},$$

so we may write

$$F(g_1(x), \ldots, g_k(x)) = \sum_{S \subseteq [n]} \widetilde{F}(S) \prod_{i \in S} \Phi(g_i(x)), \text{ where } \Phi(g_i(x)) = \frac{g_i(x) - \mathbf{E}_\mu[g_i]}{\mathbf{Var}_\mu[g_i]^{1/2}}.$$

Note that Φ normalizes g_i such that $\mathbf{E}_\mu[\Phi(g_i)] = 0$ and $\mathbf{E}_\mu[\Phi(g_i)^2] = 1$. First we claim that

$$\mathbf{H}^\mu[f^{\geq 1}] = \mathbf{H}^\mu\left[\sum_{S \neq \emptyset} \widetilde{F}(S) \prod_{i \in S} \Phi(g_i)\right] = \sum_{S \neq \emptyset} \mathbf{H}^\mu\left[\widetilde{F}(S) \prod_{i \in S} \Phi(g_i)\right].$$

It suffices to show that for any two distinct non-empty sets $S, T \subseteq [k]$, no monomial ϕ^μ_U occurs in the μ-biased spectral support of both $\widetilde{F}(S) \prod_{i \in S} \Phi(g_i)$ and $\widetilde{F}(T) \prod_{i \in T} \Phi(g_i)$. To see this recall that $\Phi(g_i)$ is balanced with respect to μ (*i.e.* $\mathbf{E}_\mu[\Phi(g_i)] = \mathbf{E}_\mu[\Phi(g_i)\phi^\mu_\emptyset] = 0$), and so every monomial ϕ^μ_U in the support of $\widetilde{F}(S) \prod_{i \in S} \Phi(g_i)$ is of the form $\prod_{i \in S} \phi^\mu_{U_i}$ where U_i is a non-empty subset of the relevant variables of g_i (*i.e.* $\{(i-1)\ell + 1, \ldots, i\ell\}$); likewise for monomials in the support of $\widetilde{F}(T) \prod_{i \in T} \Phi(g_i)$. In other words the non-empty subsets of $[k]$ induce a partition of the μ-biased Fourier support of f, where ϕ^μ_U is mapped to $\emptyset \neq S \subseteq [k]$ if and only if U contains a relevant variable of g_i for every $i \in S$ and none of the relevant variables of g_j for any $j \notin S$.

With this identity in hand we have

$$\begin{aligned}
\mathbf{H}^\mu[f^{\geq 1}] &= \sum_{S \neq \emptyset} \mathbf{H}^\mu\left[\widetilde{F}(S) \prod_{i \in S} \Phi(g_i)\right] \\
&= \sum_{S \neq \emptyset} \varphi(\widetilde{F}(S)) + \widetilde{F}(S)^2 \sum_{i \in S} \mathbf{H}^\mu[\Phi(g_i)]. \\
&= \sum_{S \neq \emptyset} \varphi(\widetilde{F}(S)) + \widetilde{F}(S)^2 \sum_{i \in S} \left(\frac{\mathbf{H}^\mu[g_i - \mathbf{E}_\mu[g_i]]}{\mathbf{Var}_\mu[g_i]} + \varphi\left(\frac{1}{\mathbf{Var}_\mu[g_i]^{1/2}}\right) \mathop{\mathbf{Var}}_\mu[g_i] \right) \\
&= \mathbf{H}^\eta[F^{\geq 1}] + \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \frac{\mathbf{H}^\mu[g_i^{\geq 1}]}{\mathbf{Var}_\mu[g_i]},
\end{aligned}$$

where the second and third equalities are two applications of Lemma 1 (for the second equality we view $\widetilde{F}(S)$ as a constant function with $\mathbf{H}^\mu[\widetilde{F}(S)] = \varphi(\widetilde{F}(S))$). By the same reasoning, we also have

$$\begin{aligned}
\mathbf{Inf}^\mu[f] \;=\; \sum_{S \neq \emptyset} \mathbf{Inf}^\mu\left[\widetilde{F}(S) \prod_{i \in S} \Phi(g_i(x^i))\right] &= \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \mathbf{Inf}^\mu[\Phi(g_i)] \\
&= \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \frac{\mathbf{Inf}^\mu[g_i]}{\mathbf{Var}_\mu[g_i]}.
\end{aligned}$$

Here the second equality is by Lemma 1, again viewing $\widetilde{F}(S)$ as a constant function with $\mathbf{Inf}^\mu[\widetilde{F}(S)] = 0$, and the third equality uses the fact that $\mathbf{Inf}^\mu[\alpha f] = \alpha^2 \cdot \mathbf{Inf}^\mu[f]$ and $\mathbf{Inf}^\mu[g_i - \mathbf{E}_\mu[g_i]] = \mathbf{Inf}^\mu[g_i]$. Finally we see that

$$\mathbf{Var}_\mu[f] = \sum_{S \neq \emptyset} \mathbf{Var}_\mu\left[\widetilde{F}(S) \prod_{i \in S} \Phi(g_i)\right] = \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \prod_{i \in S} \mathbf{Var}_\mu[\Phi(g_i)] = \sum_{S \neq \emptyset} \widetilde{F}(S)^2,$$

where the last quantity is $\mathbf{Var}_\eta[F]$. Here the second equality uses the fact that the functions $\Phi(g_i)$ are on disjoint sets of variables (and therefore statistically independent when viewed as random variables), and the third equality holds since $\mathbf{Var}_\mu[\Phi(g_i)] = \mathbf{E}[\Phi(g_i)^2] - \mathbf{E}[\Phi(g_i)]^2 = 1$. $\square$

We are now ready to prove our main theorem:

Theorem 1. *Let $F : \{-1,1\}^k \to \mathbb{R}$, and $g_1, \dots, g_k : \{-1,1\}_\mu^{k\ell} \to \{-1,1\}$ where each g_i depends only on the ℓ coordinates in $\{(i-1)\ell + 1, \dots, i\ell\}$. Let $f(x) = F(g_1(x), \dots, g_k(x))$ and suppose $C > 0$ satisfies*

1. $\mathbf{H}^\mu[g_i^{\geq 1}] \leq C \cdot (\mathbf{Inf}^\mu[g_i] - \mathbf{Var}_\mu[g_i])$ *for all $i \in [k]$.*
2. $\mathbf{H}^\eta[F^{\geq 1}] \leq C \cdot (\mathbf{Inf}^\eta[F] - \mathbf{Var}_\eta[F])$*, where $\eta = \langle \mathbf{E}_\mu[g_1], \dots, \mathbf{E}_\mu[g_k] \rangle$.*

Then $\mathbf{H}^\mu[f^{\geq 1}] \leq C \cdot (\mathbf{Inf}^\mu[f] - \mathbf{Var}_\mu[f])$.

Proof. By our first assumption each g_i satisfies $\mathbf{Inf}^\mu[g_i] \geq \frac{1}{C}\mathbf{H}^\mu[g^{\geq 1}] + \mathbf{Var}_\mu[g_i]$, and so combining this with equation (2) of Proposition 2 we have

$$\begin{aligned}\mathbf{Inf}^\mu[f] \;&=\; \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \frac{\mathbf{Inf}^\mu[g_i]}{\mathbf{Var}_\mu[g_i]} \geq \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \left(\frac{\mathbf{H}^\mu[g_i^{\geq 1}]}{C\,\mathbf{Var}_\mu[g_i]} + 1 \right) \\ &= \mathbf{Inf}^\eta[F] + \frac{1}{C} \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \frac{\mathbf{H}^\mu[g_i^{\geq 1}]}{\mathbf{Var}_\mu[g_i]} \qquad (4)\end{aligned}$$

This along with equations (1) and (3) of Proposition 2 completes the proof:

$$\begin{aligned}\mathbf{H}^\mu[f^{\geq 1}] &= \mathbf{H}^\eta[F^{\geq 1}] + \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \frac{\mathbf{H}^\mu[g_i^{\geq 1}]}{\mathbf{Var}_\mu[g_i]} \\ &\leq C \cdot (\mathbf{Inf}^\eta[F] - \mathbf{Var}_\eta[F]) + \sum_{S \neq \emptyset} \widetilde{F}(S)^2 \sum_{i \in S} \frac{\mathbf{H}^\mu[g_i^{\geq 1}]}{\mathbf{Var}_\mu[g_i]} \\ &\leq C \cdot (\mathbf{Inf}^\mu[f] - \mathbf{Var}_\eta[F]) \;=\; C \cdot (\mathbf{Inf}^\mu[f] - \mathbf{Var}_\mu[f]).\end{aligned}$$

Here the first equality is by (1), the first inequality by our second assumption, the second inequality by (4), and finally the last identity by (3). $\square$

4 Distribution-independent bound for FEI$^+$

In this section we prove that μ-biased FEI$^+$ holds for all Boolean functions $F : \{-1,1\}_\mu^k \to \{-1,1\}$ with factor C independent of the biases $\mu_1, \dots, \mu_k$ of μ. When $\mu = \langle 0, \dots 0 \rangle$ is the uniform distribution it is well-known that the FEI conjecture holds with factor $C = O(\log k)$, and a bound of $C \leq 2^k$ is trivial since $\mathbf{Inf}[F]$ is always an integer multiple of 2^{-k} and $\mathbf{H}[F] \leq 1$; neither proofs carry through to the setting of product distributions. We remark that even verifying the seemingly simple claim "there exists a universal constant C such that $\mathbf{H}^\mu[\mathsf{MAJ}_3] \leq C \cdot (\mathbf{Inf}^\mu[\mathsf{MAJ}_3] - \mathbf{Var}_\mu[\mathsf{MAJ}_3])$ for all product distributions $\mu \in [0,1]^3$", where MAJ_3 the majority function over 3 variables, turns out to be technically cumbersome.

The high-level strategy is to bound each of the $2^k - 1$ terms of $\mathbf{H}^\mu[F^{\geq 1}]$ separately; due to space considerations we defer the proof the main lemma to Appendix B.

Lemma 2. *Let $F : \{-1,1\}^k_\mu \to \{-1,1\}$. Let $S \subseteq [k]$, $S \neq \emptyset$, and suppose $\widetilde{F}(S) \neq 0$. For any $j \in S$ we have*

$$\widetilde{F}(S)^2 \log\left(\frac{\prod_{i\in S}\sigma_i^2}{\widetilde{F}(S)^2}\right) \leq \frac{2^{2k}}{\ln 2} \cdot \mathop{\mathbf{Var}}_{\mu}[D_{\phi_j^\mu} F].$$

Theorem 2. *Let $F : \{-1,1\}^k_\mu \to \{-1,1\}$. Then*

$$\mathbf{H}^\mu[F^{\geq 1}] \leq 2^{O(k)} \cdot (\mathbf{Inf}^\mu[F] - \mathbf{Var}_\mu[F]).$$

Proof. The claim can be equivalently stated as $\mathbf{H}^\mu[F^{\geq 1}] \leq 2^{O(k)} \sum_{i=1}^n \mathbf{Var}_\mu[D_{\phi_i^\mu} F]$, since

$$\sum_{i=1}^n \mathbf{Var}[D_{\phi_i^\mu} F] = \sum_{|S|\geq 2} |S| \cdot \widetilde{F}(S)^2 \leq 2 \sum_{|S|\geq 2} (|S|-1) \cdot \widetilde{F}(S)^2 = 2 \cdot (\mathbf{Inf}^\mu[F] - \mathbf{Var}_\mu[F]).$$

By Lemma 2, for every $S \neq \emptyset$ that contributes $\varphi(\widetilde{F}(S))$ to $\mathbf{H}^\mu[F^{\geq 1}]$ we have $\varphi(\widetilde{F}(S)) \leq 2^{O(k)} \mathbf{Var}_\mu[D_{\phi_j^\mu} F]$, where j is any element of S. Summing over all $2^k - 1$ non-empty subsets S of $[k]$ completes the proof. $\square$

4.1 FEI$^+$ for read-once formulas

Finally, we combine our two main results so far, the composition theorem (Theorem 1) and the distribution-independent universal bound (Theorem 2), to prove Conjecture 1 for read-once formulas with arbitrary gates of bounded arity.

Definition 1. *Let $\mathcal{B}$ be a set of Boolean functions. We say that a Boolean function f is a formula over the basis $\mathcal{B}$ if f is computable a formula with gates belonging to $\mathcal{B}$. We say that f is a read-once formula over $\mathcal{B}$ if every variable appears at most once in the formula for f.*

Corollary 1. *Let $C > 0$ and $\mathcal{B}$ be a set of Boolean functions, and suppose $\mathbf{H}^\mu[F] \leq C \cdot (\mathbf{Inf}^\mu[F] - \mathbf{Var}_\mu[F])$ for all $F \in \mathcal{B}$ and product distributions μ. Let $\mathcal{C}$ be the class of read-once formulas over the basis $\mathcal{B}$. Then $\mathbf{H}^\mu[f] \leq C \cdot (\mathbf{Inf}^\mu[f] - \mathbf{Var}_\mu[f])$ for all $f \in \mathcal{C}$ and product distributions μ.*

Proof. We proceed by structural induction on the formula computing f. The base case holds since the μ-biased Fourier expansion of the dictator x_1 and anti-dictator $-x_i$ is $\pm(\mu_1 + \sigma_1\phi_1^\mu(x))$ and so $\mathbf{H}^\mu[f^{\geq 1}] = \widetilde{f}(\{1\})^2 \log(\sigma_1^2/\widetilde{f}(\{1\})^2) = \sigma_1^2 \log(\sigma_1^2/\sigma_1^2) = 0$.

For the inductive step, suppose $f = F(g_1, \ldots, g_k)$, where $F \in \mathcal{B}$ and $g_1, \ldots, g_k$ are read-once formulas over $\mathcal{B}$ over disjoint sets of variables. Let μ be any product distribution over the domain of f. By our induction hypothesis we have $\mathbf{H}^\mu[g_i^{\geq 1}] \leq C \cdot (\mathbf{Inf}^\mu[g_i] - \mathbf{Var}_\mu[g_i])$ for all $i \in [k]$, satisfying the first requirement of Theorem 1. Next, by our assumption on $F \in \mathcal{B}$, we have $\mathbf{H}^\eta[F^{\geq 1}] \leq C \cdot (\mathbf{Inf}^\eta[F] - \mathbf{Var}_\eta[F])$ for all product distributions η, and in particular, $\eta = \langle \mathbf{E}_\mu[g_1], \ldots, \mathbf{E}_\mu[g_k] \rangle$, satisfying the second requirement of Theorem 1. Therefore, by Theorem 1 we conclude that $\mathbf{H}^\mu[f] \leq C \cdot (\mathbf{Inf}^\mu[f] - \mathbf{Var}_\mu[f])$. □

By Theorem 2, for any set $\mathcal{B}$ of Boolean functions with maximum arity k and product distribution μ, every $F \in \mathcal{B}$ satisfies $\mathbf{H}^\mu[F] \leq 2^{O(k)} \cdot (\mathbf{Inf}^\mu[F] - \mathbf{Var}_\mu[q])$. Combining this with Corollary 1 yields the following:

Theorem 3. *Let $\mathcal{B}$ be a set of Boolean functions with maximum arity k, and $\mathcal{C}$ be the class of read-once formulas over the basis $\mathcal{B}$. Then $\mathbf{H}^\mu[f] \leq 2^{O(k)} \cdot (\mathbf{Inf}^\mu[f] - \mathbf{Var}_\mu[f])$ for all $f \in \mathcal{C}$ and product distributions μ.*

5 Lower bound on the constant of the FEI conjecture

The tools we develop in this paper also yield an explicit function f achieving the best-known ratio between $\mathbf{H}[f]$ and $\mathbf{Inf}[f]$ (*i.e.* a lower bound on the constant C in the FEI conjecture). We will use the following special case of Proposition 2 on the behavior of spectral entropy and total influence under composition:

Lemma 3 (Amplification lemma). *Let $F : \{-1,1\}^k \to \{-1,1\}$ and $g : \{-1,1\}^\ell \to \{-1,1\}$ be balanced Boolean functions. Let $f_0 = g$, and for all $m \geq 1$, define $f_m = F(f_{m-1}(x^1), \ldots, f_{m-1}(x^k))$. Then*

$$\mathbf{H}[f_m] = \mathbf{H}[g] \cdot \mathbf{Inf}[F]^m + \mathbf{H}[F] \cdot \frac{\mathbf{Inf}[F]^m - 1}{\mathbf{Inf}[F] - 1}$$
$$\mathbf{Inf}[f_m] = \mathbf{Inf}[g] \cdot \mathbf{Inf}[F]^m.$$

In particular, if $F = g$ we have

$$\frac{\mathbf{H}[f_m]}{\mathbf{Inf}[f_m]} = \frac{\mathbf{H}[F]}{\mathbf{Inf}[F]} + \frac{\mathbf{H}[F]}{\mathbf{Inf}[F](\mathbf{Inf}[F]-1)} - \frac{\mathbf{H}[F]}{\mathbf{Inf}[F]^{m+1}(\mathbf{Inf}[F]-1)}.$$

Proof. Since the composition of a balanced function with another remains balanced, we have the recurrence relations $\mathbf{H}[f_m] = \mathbf{H}[f_{m-1}] \cdot \mathbf{Inf}[F] + \mathbf{H}[F]$ and $\mathbf{H}[f_m] = \mathbf{H}[f_{m-1}] \cdot \mathbf{Inf}[F] + \mathbf{H}[F]$ as special cases of Proposition 2. Solving them yields the claim. □

Theorem 4. *There exists an infinite family of functions $f_m : \{-1,1\}^{6^m} \to \{-1,1\}$ such that $\lim_{m\to\infty} \mathbf{H}[f_m]/\mathbf{Inf}[f_m] \geq 6.278944$.*

Proof. Let

$$g = (\overline{x}_1 \wedge x_2 \wedge x_3) \vee (x_1 \wedge \overline{x}_2 \wedge x_4) \vee (x_1 \wedge \overline{x}_2 \wedge x_5 \wedge x_6) \vee (x_1 \wedge x_2 \wedge x_3) \vee (x_1 \wedge x_2 \wedge x_4 \wedge x_5).$$

It can be checked that g is a balanced function with $\mathbf{H}[F] \geq 3.92434$ and $\mathbf{Inf}[F] = 1.625$. Applying Lemma 3 with $F = g$, we get

$$\lim_{m\to\infty} \frac{\mathbf{H}[f_m]}{\mathbf{Inf}[f_m]} \geq \frac{3.92434}{1.625} + \frac{3.92434}{1.625 \times 0.625} = 6.278944.$$

□

References

1. Friedgut, E., Kalai, G.: Every monotone graph property has a sharp threshold. Proceedings of the American Mathematical Society **124**(10) (1996) 2993–3002
2. O'Donnell, R., Wright, J., Zhou, Y.: The Fourier Entropy-Influence Conjecture for certain classes of boolean functions. In: Proceedings of the 38th Annual International Colloquium on Automata, Languages and Programming. (2011) 330–341
3. Kalai, G.: The entropy/influence conjecture. Posted on Terence Tao's *What's new* blog, http://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/ (2007)
4. Kahn, J., Kalai, G., Linial, N.: The influence of variables on Boolean functions. In: Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science. (1988) 68–80
5. Mansour, Y.: Learning Boolean functions via the Fourier Transform. In Roychowdhury, V., Siu, K.Y., Orlitsky, A., eds.: Theoretical Advances in Neural Computation and Learning. Kluwer Academic Publishers (1994) 391–424
6. Gopalan, P., Kalai, A., Klivans, A.: Agnostically learning decision trees. In: Proceedings of the 40th Annual ACM Symposium on Theory of Computing. (2008) 527–536
7. Gopalan, P., Kalai, A., Klivans, A.: A query algorithm for agnostically learning DNF? In: Proceedings of the 21st Annual Conference on Learning Theory. (2008) 515–516
8. De, A., Etesami, O., Trevisan, L., Tulsiani, M.: Improved pseudorandom generators for depth 2 circuits. In: Proceedings of the 14th Annual International Workshop on Randomized Techniques in Computation. (2010) 504–517
9. Friedgut, E.: Boolean functions with low average sensitivity depend on few coordinates. Combinatorica **18**(1) (1998) 27–36
10. Servedio, R.: On learning monotone DNF under product distributions. Information and Computation **193**(1) (2004) 57–74
11. O'Donnell, R., Servedio, R.: Learning monotone decision trees in polynomial time. SIAM Journal on Computing **37**(3) (2008) 827–844
12. Klivans, A., Lee, H., Wan, A.: Mansour's Conjecture is true for random DNF formulas. In: Proceedings of the 23rd Annual Conference on Learning Theory. (2010) 368–380
13. Heiman, R., Newman, I., Wigderson, A.: On read-once threshold formulae and their randomized decision in tree complexity. Theor. Comput. Sci. **107**(1) (1993) 63–76
14. Buhrman, H., de Wolf, R.: Complexity measures and decision tree complexity: a survey. Theoretical Computer Science **288**(1) (2002) 21–43
15. Keller, N., Mossel, E., Schlank, T.: A note on the entropy/influence conjecture. Discrete Mathematics **312**(22) (2012) 3364–3372
16. Bourgain, J., Kahn, J., Kalai, G., Katznelson, Y., Linial, N.: The influence of variables in product spaces. Israel Journal of Mathematics **77**(1) (1992) 55–64

A Biased Fourier Analysis

Theorem 5 (Fourier expansion). *Let $\mu = \langle \mu_1, \ldots, \mu_n \rangle$ be a sequence of biases. The μ-biased Fourier expansion of $f : \{-1,1\}^n \to \mathbb{R}$ is*

$$f(x) = \sum_{S \subseteq [n]} \widetilde{f}(S) \phi_S^{\mu}(x),$$

where

$$\phi_S^{\mu}(x) = \prod_{i \in S} \frac{x_i - \mu_i}{\sigma_i} \quad \text{and} \quad \widetilde{f}(S) = \mathop{\mathbf{E}}_{\mu}[f(\boldsymbol{x}) \phi_S^{\mu}(\boldsymbol{x})],$$

and $\sigma_i^2 = \mathbf{Var}_{\mu}[x_i] = 1 - \mu_i^2$.

The μ-biased spectral support of f is the collection $\mathcal{S} \subseteq 2^{[n]}$ of subsets $S \subseteq [n]$ such that $\widetilde{f}(S) \neq 0$. We write $f^{\geq k}$ to denote $\sum_{|S| \geq k} \widetilde{f}(S) \phi_S^{\mu}(x)$, the projection of f onto its monomials of degree at least k.

Theorem 6 (Parseval's identity). *Let $f : \{-1,1\}_{\mu}^n \to \mathbb{R}$. Then $\sum_{S \subseteq [n]} \widetilde{f}(S)^2 = \mathbf{E}_{\mu}[f(\boldsymbol{x})^2]$. In particular, if the range of f is $\{-1,1\}$ then $\sum_{S \subseteq [n]} \widetilde{f}(S)^2 = 1$.*

Definition 2 (Influence). *Let $f : \{-1,1\}_{\mu}^n \to \mathbb{R}$. The influence of variable $i \in [n]$ on f is $\mathbf{Inf}_i^{\mu}[f] = \mathbf{E}_{\rho}[\mathbf{Var}_{\mu_i}[f_{\rho}]]$, where ρ is a μ-biased random restriction to the coordinates in $[n] \backslash \{i\}$. The total influence of f, denoted $\mathbf{Inf}^{\mu}[f]$, is $\sum_{i=1}^n \mathbf{Inf}_i^{\mu}[f]$.*

We recall a few basic Fourier formulas. The expectation of f is given by $\mathbf{E}_{\mu}[f] = \widetilde{f}(\emptyset)$ and its variance $\mathbf{Var}_{\mu}[f] = \sum_{S \neq \emptyset} \widetilde{f}(S)^2$. For each $i \in [n]$, $\mathbf{Inf}_i^{\mu}[f] = \sum_{S \ni i} \widetilde{f}(S)^2$ and so $\mathbf{Inf}^{\mu}[f] = \sum_{S \subseteq [n]} |S| \cdot \widetilde{f}(S)^2$. We omit the sub- and superscripts when $\mu = \langle 0, \ldots, 0 \rangle$ is the uniform distribution. Comparing the Fourier formulas for variance and total influence yields the Poincaré inequality for functions $f : \{-1,1\}_{\mu}^n \to \mathbb{R}$:

Theorem 7 (Poincaré inequality). *Let $f : \{-1,1\}_{\mu}^n \to \mathbb{R}$. Then $\mathbf{Inf}^{\mu}[f] \leq \mathbf{Var}_{\mu}[f]$.*

Recall that the i-th discrete derivative operator for $f : \{-1,1\}^n \to \{-1,1\}$ is defined to be

$$D_{x_i}(x) = \tfrac{1}{2} \left(f(x^{i \leftarrow 1}) - f(x^{i \leftarrow -1}) \right),$$

and for $S \subseteq [n]$ we write $D_{x^S} f$ to denote $\circ_{i \in S} D_{x_i} f$.

Definition 3 (Discrete derivative). *The i-th discrete derivative operator $D_{\phi_i^{\mu}}$ with respect to the μ-biased product distribution on $\{-1,1\}^n$ is defined by $D_{\phi_i^{\mu}} f(x) = \sigma_i D_{x_i} f(x)$.*

With respect to the μ-biased Fourier expansion of $f : \{-1,1\}_{\mu}^n \to \mathbb{R}$ the operator $D_{\phi_i^{\mu}}$ satisfies

$$D_{\phi_i^{\mu}} f = \sum_{S \ni i} \widetilde{f}(S) \phi_S^{\mu},$$

and so for any $S \subseteq [n]$ we have $\widetilde{f}(S) = \mathbf{E}[\circ_{i \in S} D_{\phi_i^{\mu}} f] = \prod_{i \in S} \sigma_i \, \mathbf{E}_{\mu}[(D_{x^S} f)]$.

B Omitted Proofs

Lemma 1. *Let $\Phi_1, \ldots, \Phi_k : \{-1,1\}^{k\ell}_\mu \to \mathbb{R}$ where each Φ_i depends only on the ℓ coordinates in $\{(i-1)\ell+1, \ldots, i\ell\}$. Then*

$$\mathbf{H}^\mu[\Phi_1 \cdots \Phi_k] = \sum_{i=1}^k \mathbf{H}^\mu[\Phi_i] \prod_{j \neq i} \mathop{\mathbf{E}}_\mu[\Phi_j^2] \text{ and } \mathbf{Inf}^\mu[\Phi_1 \cdots \Phi_k] = \sum_{i=1}^k \mathbf{Inf}^\mu[\Phi_i] \prod_{j \neq i} \mathop{\mathbf{E}}_\mu[\Phi_j^2].$$

Proof. We prove both formulas by induction on k, noting that the bases cases are trivially true. For the inductive step, we define $h(x) = \prod_{i \in [k-1]} \Phi_i(x)$ and see that

$$\begin{aligned}
\mathbf{H}^\mu[h \cdot \Phi_k] &= \sum_{\substack{S \subseteq [(k-1)\ell] \\ T \subseteq \{(k-1)\ell+1, \ldots k\ell\}}} \widetilde{h}(S)^2 \widetilde{\Phi_k}(T)^2 \log\left(\frac{\prod_{i \in S \cup T} \sigma_i^2}{\widetilde{h}(S)^2 \widetilde{\Phi_k}(T)^2}\right) \\
&= \sum_{S,T} \widetilde{h}(S)^2 \widetilde{\Phi_k}(T)^2 \left[\log\left(\frac{\prod_{i \in S} \sigma_i^2}{\widetilde{h}(S)^2}\right) + \log\left(\frac{\prod_{i \in T} \sigma_i^2}{\widetilde{\Phi_k}(T)^2}\right)\right] \\
&= \mathop{\mathbf{E}}_\mu[h^2] \cdot \mathbf{H}^\mu[\Phi_k] + \mathop{\mathbf{E}}_\mu[\Phi_k^2] \cdot \mathbf{H}^\mu[h] \\
&= \prod_{i \in [k-1]} \mathop{\mathbf{E}}_\mu[\Phi_i^2] \cdot \mathbf{H}^\mu[\Phi_k] + \mathop{\mathbf{E}}_\mu[\Phi_k^2] \left(\sum_{i=1}^{k-1} \mathbf{H}^\mu[\Phi_i] \prod_{j \neq i} \mathop{\mathbf{E}}_\mu[\Phi_j^2]\right) \\
&= \sum_{i=1}^k \mathbf{H}^\mu[\Phi_i] \prod_{j \neq i} \mathop{\mathbf{E}}_\mu[\Phi_j^2].
\end{aligned}$$

Here in the first equality we use the fact that if $f : \{-1,1\}^n_\mu \to \mathbb{R}$ does not depend on coordinate $i \in [n]$ then $\widetilde{f}(S) = 0$ for all $S \ni i$ (*i.e.* the Fourier spectrum of f is supported on sets containing only its relevant variables). The third equality is by Parseval's, and the fourth by the induction hypothesis applied to h.

The formula for influence follows from a similar derivation:

$$\begin{aligned}
\mathbf{Inf}^\mu[h \cdot \Phi_k] &= \sum_{\substack{S \subseteq [(k-1)\ell] \\ T \subseteq \{(k-1)\ell+1, \ldots k\ell\}}} |S \cup T| \cdot \widetilde{h}(S)^2 \widetilde{\Phi_k}(T)^2 \\
&= \sum_{S,T} |T| \cdot \widetilde{h}(S)^2 \widetilde{\Phi_k}(T)^2 + \sum_{S,T} |S| \cdot \widetilde{h}(S)^2 \widetilde{\Phi_k}(T)^2 \\
&= \mathop{\mathbf{E}}_\mu[h^2] \cdot \mathbf{Inf}^\mu[\Phi_k] + \mathop{\mathbf{E}}_\mu[\Phi_k^2] \cdot \mathbf{Inf}^\mu[h] \\
&= \prod_{i \in [k-1]} \mathop{\mathbf{E}}_\mu[\Phi_i^2] \cdot \mathbf{Inf}^\mu[\Phi_k] + \mathop{\mathbf{E}}_\mu[\Phi_k^2] \left(\sum_{i=1}^{k-1} \mathbf{Inf}^\mu[\Phi_i] \prod_{j \neq i} \mathop{\mathbf{E}}_\mu[\Phi_j^2]\right) \\
&= \sum_{i=1}^k \mathbf{Inf}^\mu[\Phi_i] \prod_{j \neq i} \mathop{\mathbf{E}}_\mu[\Phi_j^2],
\end{aligned}$$

and this completes the proof. $\square$

Lemma 2. *Let $F : \{-1,1\}^k_\mu \to \{-1,1\}$. Let $S \subseteq [k]$, $S \neq \emptyset$, and suppose $\widetilde{F}(S) \neq 0$. For any $j \in S$ we have*

$$\widetilde{F}(S)^2 \log\left(\frac{\prod_{i\in S}\sigma_i^2}{\widetilde{F}(S)^2}\right) \leq \frac{2^{2k}}{\ln 2}\cdot \mathop{\mathbf{Var}}_{\mu}[D_{\phi_j^\mu}F].$$

Proof. Recall that $\widetilde{F}(S) = \mathbf{E}_\mu[\circ_{i\in S} D_{\phi_i^\mu} f] = \prod_{i\in S}\sigma_i\,\mathbf{E}_\mu[D_{x^S} f]$, and so

$$\begin{aligned}
\widetilde{F}(S)^2 \log\left(\frac{\prod_{i\in S}\sigma_i^2}{\widetilde{F}(S)^2}\right) &= \prod_{i\in S}\sigma_i^2 \cdot \mathop{\mathbf{E}}_{\mu}[D_{x^S}F]^2 \log\left(\frac{1}{\mathbf{E}[D_{x^S}F]^2}\right)\\
&\leq \frac{1}{\ln 2}\prod_{i\in S}\sigma_i^2 \cdot \big|\mathop{\mathbf{E}}_{\mu}[D_{x^S}F]\big|\\
&\leq \frac{1}{\ln 2}\prod_{i\in S}\sigma_i^2 \mathop{\mathbf{Pr}}_{\mu}[D_{x^S}F \neq 0].
\end{aligned}$$

Here the first inequality holds since $t^2\log(1/t^2) \leq t/\ln(2)$ for all $t \in \mathbb{R}^+$, and the second uses the fact that $D_{x^S}F$ is bounded within $[-1,1]$. Therefore it suffices to argue that

$$\begin{aligned}
\prod_{i\in S}\sigma_i^2 \mathop{\mathbf{Pr}}_{\mu}[D_{x^S}F \neq 0] &\leq 2^{2k}\cdot \mathop{\mathbf{Var}}_{\mu}[D_{\phi_j^\mu}F]\\
&= 2^{2k}\sigma_j^2 \cdot \mathop{\mathbf{Var}}_{\mu}[D_jF]\\
&= 2^{2k}\sigma_j^2 \mathop{\mathbf{E}}_{y\in\{-1,1\}^{[n]\setminus S}}\left[\mathop{\mathbf{E}}_{z\in\{-1,1\}^{S\setminus\{j\}}}\left[((D_jF)|_y(z)-\mu)^2\right]\right],
\end{aligned}$$

where $\mu = \mathbf{E}[D_jF]$ and $(D_jF)|_y$ denotes the restriction of D_jF where the coordinates in $[n]\backslash S$ are set according to y. We first rewrite the desired inequality above as

$$\left(2^{-2k}\prod_{i\in S\setminus\{j\}}\sigma_i^2\right)\mathop{\mathbf{E}}_{y\in\{-1,1\}^{[n]\setminus S}}[\mathbf{1}_{D_{x^S}F(y)\neq 0}]$$
$$\leq \mathop{\mathbf{E}}_{y\in\{-1,1\}^{[n]\setminus S}}\left[\mathop{\mathbf{E}}_{z\in\{-1,1\}^{S\setminus\{j\}}}\left[((D_jF)|_y(z)-\mu)^2\right]\right]$$

and argue that this holds point-wise: for every $y \in [n]\backslash S$ such that $D_{x^S}F(y) \neq 0$,

$$\mathbf{E}\left[((D_jF)|_y(z)-\mu)^2\right] \geq 2^{-2k}\prod_{i\in S\setminus\{j\}}\sigma_i^2.$$

To see this, fix $y \in \{-1,1\}^{[n]\setminus S}$ such that $(D_{x^S}F)(y) \neq 0$. Viewing $(D_{x^S}F)$ as $(D_{x^{S\setminus\{j\}}}D_jF)$, it follows that $(D_jF)|_y$ is non-constant. Since $(D_jF)|_y$ takes values in $\{-1,0,1\}$, there must exist some $z^* \in \{-1,1\}^{S\setminus\{j\}}$ such that $|(D_jF)|_y(z^*)-$

$\mu| \geq \frac{1}{2}$ and so indeed

$$\begin{aligned}\mathbf{E}\left[((D_jF)|_y(z)-\mu)^2\right] &\geq \left(\frac{1}{2}\right)^2 \mathbf{Pr}[z=z^*] \\ &= \frac{1}{4}\prod_{i\in S\setminus\{j\}} \frac{1\pm\mu_i}{2} \;\geq\; \frac{1}{4}\prod_{i\in S\setminus\{j\}} \frac{\sigma_i^2}{4} \;\geq\; 2^{-2k}\prod_{i\in S\setminus\{j\}} \sigma_i^2.\end{aligned}$$

$\square$