On the Fourier Entropy Influence Conjecture

Satya Lokam

Microsoft Research
Bangalore

19 December 2015

What is this Conjecture?

- For $f : \{0,1\}^n \to \{+1,-1\}$, its Fourier coefficients are denoted $\{\hat{f}(S) : S \subseteq [n]\}$.
- $\sum_S \widehat{f}^2(S) = 1$. So, $\widehat{f}^2(\cdot)$ defines a distribution on $\{S \ : \ S \subseteq [n]\}$.
- The (Shannon) entropy of this distribution is the *Fourier Entropy of f*:

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

- The *Influence of f*, $\mathsf{Inf}(f)$, is the expected number of coordinates of a random input which, when flipped, will cause the value of f to be changed.
- **Fourier Entropy Influence Conjecture** (Friedgut-Kalai, 1996) :
  There exists a universal constant C such that for all $f : \{0,1\}^n \to \{+1,-1\}$,

$$\mathbb{H}(f) \leqslant C \cdot \mathsf{Inf}(f).$$

Outline

Outline

Fourier Transforms of Boolean Functions

- Vector space of all $f : \{0,1\}^n \longrightarrow \mathbb{R}$; Inner Product: $\langle f, g \rangle := 2^{-n} \sum_{x \in \{0,1\}^n} f(x)g(x)$
- Orthonormal basis of characters: $\chi_S(x) := (-1)^{\sum_{i \in S} x_i}$ for $S \subseteq [n]$ – Parity on S
- Fourier Coefficient: $\hat{f}(S) = \langle f, \chi_S \rangle = 2^{-n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x)$
  – Correlation with Parity on S
- Fourier expansion: $f(x) = \sum_S \hat{f}(S)\chi_S(x)$
- Norm: $\|f\| = \sqrt{\langle f, f \rangle} = \mathbb{E}_x[f(x)^2]$
- Parseval: $\|f\|^2 = \sum_S \widehat{f}^2(S)$
- For Boolean $f : \{0,1\}^n \rightarrow \{+1,-1\}$, $\sum_S \widehat{f}^2(S) = \|f\|^2 = 1$

Influence and Sensitivity of Boolean Functions

- The *influence of f in the i-th direction*:

$$\mathsf{Inf}_i(f) = \Pr[x \in \{0,1\}^n \ : \ f(x) \neq f(x \oplus e_i)] \ ,$$

  where $x \oplus e_i$ is obtained from x by flipping the i-th bit of x.
- The (total) influence of f: $\mathsf{Inf}(f) = \sum_{i=1}^{n} \mathsf{Inf}_i(f)$.
- Kahn-Kalai-Linial – KKL88: $\mathsf{Inf}_i(f) = \sum_{S \ni i} \hat{f}(S)^2$ and hence $\mathsf{Inf}(f) = \sum_{S \subseteq [n]} |S| \hat{f}(S)^2$
- For $x \in \{0,1\}^n$, the *sensitivity of f at x*: $\mathsf{s}_f(x) := |\{i : f(x) \neq f(x \oplus e_i), 1 \leqslant i \leqslant n\}|$,
- The average sensitivity of f: $\mathsf{as}(f) := 2^{-n} \sum_{x \in \{0,1\}^n} s_f(x)$.
- Easy: $\mathsf{Inf}(f) = \mathsf{as}(f)$ and hence $\mathsf{as}(f) = \sum_{S \subseteq [n]} |S| \hat{f}(S)^2$.

Fourier-Entropy Influence (FEI) Conjecture

Friedgut and Kalai 1996:
There exists a universal constant C such that for all $f : \{0,1\}^n \to \{+1,-1\}$,

$$\sum_{S \subseteq [n]} \widehat{f}^2(S) \log \frac{1}{\widehat{f}^2(S)} \leqslant C \cdot \mathsf{as}(f) = C \cdot \sum_{S \subseteq [n]} |S| \hat{f}(S)^2 .$$

If the spectrum of a Boolean function appears “smeared,” then its total influence must be large, it must spread well into “high” degree coefficients.

Outline

Why prove this Conjecture?

- Sharp Thresholds for monotone random graph properties
- Implies KKL theorem
- Implies Mansour's Conjecture: A DNF with m terms can be well-approximated by $m^{O(1)}$ Fourier coefficients.
  - Agnostic Learning of DNF's
  - PRG's for depth-2 circuits

Outline

FEI holds for Symmetric Functions

Theorem (O'Donnell, Wright, and Zhou, 2011)

If $f : \{0,1\}^n \to \{+1,-1\}$ is a symmetric Boolean function, i.e., $f(x) = f(\sigma(x))$ for any permutation σ on $[n]$, the $\mathbb{H}(f) \leqslant C\, \mathsf{Inf}(f)$ for a universal constant C.

Generalizes to d-part symmetric f, i.e., f is invariant under $S_{n_1} \times \cdots \times S_{n_d}$, where d is a constant.

Splitting the Entropy

$$\mathbb{H}(f) = \sum_S \widehat{f}^2(S) \log \frac{1}{\widehat{f}^2(S)}$$

Define $\mathbf{W_k}(f) := \sum_{|S|=k} \widehat{f}^2(S)$

$$= \sum_{k=0}^{n} \mathbf{W_k}(f) \sum_{|S|=k} \frac{\widehat{f}^2(S)}{\mathbf{W_k}(f)} \log \frac{\mathbf{W_k}(f)}{\widehat{f}^2(S)} + \sum_{k=0}^{n} \mathbf{W_k}(f) \log \frac{1}{\mathbf{W_k}(f)}$$

$$= \sum_{k=0}^{n} \mathbf{W_k}(f) \mathbb{H}(f_k) + \mathbb{H}(\mathbf{W}(f))$$ where f_k is $\widehat{f}^2(.)$ restricted and normalized to k-sets

= Expected Level-wise Entropy + Entropy Across Levels

Entropy Across Levels

$$
\begin{aligned}
&\mathbb{H}(\mathbf{W}(f)) \\
&= \sum_{k=0}^{n} \mathbf{W_k}(f) \log \frac{1}{\mathbf{W_k}(f)} \\
&= (1-\mathbf{W_0}(f)) \sum_{k=1}^{n} \frac{\mathbf{W_k}(f)}{(1-\mathbf{W_0}(f))} \log \frac{(1-\mathbf{W_0}(f))}{\mathbf{W_k}(f)} + \mathrm{H}(\mathbf{W_0}(f)) \\
&= (1-\mathbf{W_0}(f))\, \mathbb{H}(\mathbf{W_1}(f), \ldots, \mathbf{W_n}(f)) \; + \; \mathrm{H}(\mathbf{W_0}(f))
\end{aligned}
$$

Let $p := \Pr_x[f(x) = -1]$ *and* $q = 1 - p$. *Then* $\mathbf{W_0}(f) = 1 - 4pq = 1 - \mathbf{Var}(f)$.

Entropy Across Levels

Lemma

Let X be a positive integer r.v. Then $\mathbb{H}(X) \leqslant \mathbb{E}[X]$.

It follows that $\mathbb{H}(\mathbf{W_1}(f), \ldots, \mathbf{W_n}(f)) \leqslant \mathbb{E}_W[k] = \sum_{k=1}^{n} k \cdot \mathbf{W_k}(f) = \mathsf{Inf}(f)$.

Lemma

By the isoperimetric inequality for the Boolean cube, $\mathrm{H}(4pq) \leqslant 2\,\mathsf{Inf}(f)$.

Theorem

For any $f : \{0,1\}^n \to \{+1,-1\}$, $\mathbb{H}(\mathbf{W}(f)) \leqslant 3\,\mathsf{Inf}(f)$.

Level-wise Entropy

$$\sum_{k=0}^{n} \mathbf{W_k}(f)\mathbb{H}(f_k)$$

$$\leqslant \sum_{k=0}^{n} \mathbf{W_k}(f) \log \binom{n}{k} \qquad = \text{for symmetric } f.$$

$$\leqslant \sum_{k=0}^{n} \mathbf{W_k}(f)(k \log \mathrm{e} + k \log \frac{n}{k})$$

$$= (\log \mathrm{e}) \ \mathsf{Inf}(f) + \sum_{k=0}^{n} \mathbf{W_k}(f) k \log \frac{n}{k}$$

Immediately implies $\mathbb{H}(f) = O(\mathsf{Inf}(f) \log n)$.

Level-wise Entropy

- Let $g_i = D_i f$ so $\mathrm{Inf}_i(f) = \mathbb{E}[g_i^2]$.
- For *symmetric* f, all Inf_i are equal. Let $g := D_n f$.
- Then, $\mathrm{Inf}(f) = n\,\mathbb{E}[g^2]$ and $k \cdot \mathbf{W_k}(f) = n \cdot \mathbf{W_{k-1}}(g)$.

$$\sum_{k=1}^{n} \mathbf{W_k}(f) k \ln \frac{n}{k} = \sum_{k=1}^{n} \mathbf{W_{k-1}}(g) n \ln \frac{n}{k}$$

$$= O(n) \sum_{k=1}^{n} \mathbf{W_{k-1}}(g) \left(\sum_{j=k}^{n} \frac{1}{j} \right) \quad \text{since } \ln m \approx \sum_{j=0}^{m} \frac{1}{j}.$$

$$= O(n) \sum_{j=1}^{n} \frac{1}{j} \left(\sum_{k=0}^{j-1} \mathbf{W_k}(g) \right)$$

Noise Stability

- (x, y) is a ρ-correlated pair if (i) x is uniformly random in $\{0,1\}^n$ and (ii) for each i independently, $\Pr[y_i = x_i] = (1+\rho)/2$ and $\Pr[y_i \neq x_i] = (1-\rho)/2$.
- Noise Stability of f with noise parameter ρ:

$$\mathbf{Stab}_\rho(f) = \mathbb{E}_{(x,y)\rho-\text{correlated}}[f(x)f(y)].$$

- Fourier expression: $\mathbf{Stab}_\rho(f) = \sum_S \rho^{|S|} \widehat{f}^2(S)$.

Level-wise Entropy

$$\mathbf{Stab}_\delta(g) = \sum_S \delta^{|S|}\widehat{g}^2(S) = \sum_{k=0}^{n} \delta^k \mathbf{W}_\mathbf{k}(g) \geqslant \delta^{j-1}\sum_{k=0}^{j-1} \mathbf{W}_\mathbf{k}(g).$$

With $\delta = 1 - \frac{1}{2j}$, we thus get

$$\sum_{k=0}^{j-1} \mathbf{W}_\mathbf{k}(g) \leqslant \left(1 - \frac{1}{2j}\right)^{-j+1} \mathbf{Stab}_{1-1/2j}(g) \leqslant \mathrm{e} \cdot \mathbf{Stab}_{1-1/2j}(g).$$

Lemma

For $g = D_n f$, where f is symmetric, $\mathbf{Stab}_{1-\theta/n}(g) \leqslant (2/\sqrt{\pi}) \cdot (1/\sqrt{\theta}) \cdot \mathbb{E}[g^2]$.

Level-wise Entropy

$$\sum_{j=1}^{n} \frac{1}{j} \left(\sum_{k=0}^{j-1} \mathbf{W_k}(g) \right)$$

$$\leqslant \sum_{j=1}^{n} \frac{1}{j} (2/\sqrt{\pi}) \cdot (\sqrt{2j}/\sqrt{n}) \cdot \mathbb{E}[g^2]$$

$$\leqslant c/\sqrt{n} \cdot \mathbb{E}[g^2] \sum_{j=1}^{n} \frac{1}{\sqrt{j}}$$

$$\leqslant c\ \mathbb{E}[g^2] \quad \text{using} \sum_{j=1}^{n} \frac{1}{\sqrt{j}} \leqslant 2\sqrt{n}.$$

Summarizing the proof for symmetric f

- Show entropy across levels is at most $\mathrm{Inf}(f)$ – applies to all functions
- Relate the expected level-wise entropy to its expectation w.r.t. Fourier mass on *levels of Discrete Derivatives of* f
- Reduce to bounding $\sum_{k=0}^{j-1} \frac{1}{j} \mathbf{W}_{<\mathbf{j}}(D_n f)$ for $1 \leqslant j \leqslant n$
- Relate to *Noise Stability* and bound $\sum_{k=0}^{j-1} \frac{1}{j} \mathbf{Stab}_{1-1/2j}(D_n f)$
- Bound on $\mathbf{Stab}_\delta(g)$ when g is *symmetric*

Outline

Parity composition preserves FEI Inequality

Lemma

Let $f = g_1 \oplus g_2$ *for* $g_i : \{0,1\}^{V_i} \to \{-1,+1\}$, *where* $[n] = V_1 \dot{\cup} V_2$. *Then,*

- $\mathbb{H}(f) = \mathbb{H}(g_1) + \mathbb{H}(g_2)$
- $\mathsf{Inf}(f) = \mathsf{Inf}(g_1) + \mathsf{Inf}(g_2)$

Proof: If $S = S_1 \dot{\cup} S_2$, $S_i \subseteq V_i$, $\widehat{f}(S) = \widehat{g_1}(S_1) \cdot \widehat{g_2}(S_2)$. □

It follows that $\mathbb{H}(f^{\oplus t}) = t \cdot \mathbb{H}(f)$ and $\mathsf{Inf}(f^{\oplus t}) = t \cdot \mathsf{Inf}(f)$.

Corollary (FEI inequality tensorizes under parity composition.)

If, for all $f : \{0,1\}^n \to \{+1,-1\}$, $\mathbb{H}(f) \leqslant C \cdot \mathsf{Inf}(f) + o(n)$, *then the FEI conjecture holds.*

$\{+1,-1\}$ vs. $\{0,1\}$

- For $f:\{0,1\}^n \to \{+1,-1\}$, let $f_{\mathbb{B}}$ denote its 0-1 counterpart: $f_{\mathbb{B}} \equiv \frac{1-f}{2}$.
- Let $p = \Pr[f_{\mathbb{B}} = 1] = \widehat{f_{\mathbb{B}}}(\emptyset)$, $q := 1-p$. Note $\mathbf{Var}(f_{\mathbb{B}}) = pq = \sum_{S\neq\emptyset} \widehat{f_{\mathbb{B}}}^2(S)$.
- Define

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

- To translate between $\mathbb{H}(f)$ and $\mathbf{H}(f_{\mathbb{B}})$:

$$\mathbb{H}(f) = 4 \cdot \mathbf{H}(f_{\mathbb{B}}) + \varphi(p), \quad \text{where} \tag{2}$$

$$\varphi(p) := \mathrm{H}(4pq) - 4p(\mathrm{H}(p) - \log p). \tag{3}$$

- Note $\varphi(p) \leqslant \mathrm{H}(4pq) \leqslant \mathsf{Inf}(f)$.

$\{0,1\}$-version of FEI Inequality

$$\textbf{FEI01 Inequality:} \quad \mathbf{H}(f_{\mathbb{B}}) \leqslant c \cdot \mathsf{Inf}(f) + \psi(p), \tag{4}$$

where c is a constant to be fixed later and

$$\psi(p) := p^2 \log \frac{1}{p^2} - 2\,\mathrm{H}(p). \tag{5}$$

Note that $\psi(p) \leqslant \mathsf{Inf}(f)$.

AND Composition preserves FEI01 Inequality

- Let $f = \text{AND}(g_1, g_2)$ with $g_i : \{0,1\}^{V_i} \to \{-1,+1\}$, and $V = V_1 \dot{\cup} V_2$.
- Obvious: $f_{\mathbb{B}} \equiv g_{1\mathbb{B}} \cdot g_{2\mathbb{B}}$.

Lemma

- *For all* $S \subseteq V$, $\widehat{f_{\mathbb{B}}}(S) = \widehat{g_{1\mathbb{B}}}(S \cap V_1) \cdot \widehat{g_{2\mathbb{B}}}(S \cap V_2)$
- $\mathbf{H}(f_{\mathbb{B}}) = p_2 \cdot \mathbf{H}(g_{1\mathbb{B}}) + p_1 \cdot \mathbf{H}(g_{2\mathbb{B}})$
- $\text{Inf}(f) = p_2 \cdot \text{Inf}(g_1) + p_1 \cdot \text{Inf}(g_2)$
- *For* $p_1, p_2 \in [0,1]$, $p_1 \cdot \psi(p_2) + p_2 \cdot \psi(p_1) \leqslant \psi(p_1 p_2)$.

Lemma (AND composition preserves FEI01 inequality)

Suppose $f_{\mathbb{B}} = \text{AND}(g_{1\mathbb{B}}, g_{2\mathbb{B}})$, *where the* g_i *depend on disjoint sets of variables. If each of the* g_i *satisfies the FEI01 Inequality* (4), *then so does* f.

FEI01 is preserved by NOT and OR composition

Lemma

If $f_{\mathbb{B}}$ satisfies FEI01 inequality (4), then so does $1 - f_{\mathbb{B}}$.

Proof.

$$\mathbf{H}(1 - f_{\mathbb{B}}) - \mathbf{H}(f_{\mathbb{B}}) = \psi(q) - \psi(p) = -p^2 \log \frac{1}{p^2} + q^2 \log \frac{1}{q^2}.$$

□

Corollary

Suppose $f_{\mathbb{B}} = \mathrm{OR}(g_{1\mathbb{B}}, g_{2\mathbb{B}})$, where the g_i depend on disjoint sets of variables. If each of the g_i satisfies the FEI01 Inequality (4), then so does f.

Proof.

$$1 - f_{\mathbb{B}} = (1 - g_{1\mathbb{B}}) \cdot (1 - g_{2\mathbb{B}}).$$

□

FEI01 holds for Read-Once De Morgan Formulas

Theorem (CKLS '15)

The FEI01 inequality (4) *holds for all read-once Boolean formulas using* AND, OR, *and* NOT *gates, with constant* $c = 5/2$.

Can be extended to include XOR gates.

Theorem (CKLS '15)

If f *is computed by a read-once formula using* AND, OR, XOR, *and* NOT *gates, then* $\mathbb{H}(f) \leqslant 10\,\mathsf{Inf}(f)$.

O'Donnell-Tan '15: FEI for Read-Once formulas with *arbitrary* bounded arity gates

- Consider μ-biased Fourier transform for a product distribution μ.
- Generalize the FEI statement to FEI$_\mu$.
- Informal Theorem: Given $f = h(g_1, \ldots, g_l)$, where g_i are defined on disjoint sets of variables, and each g_i satisfies FEI$_{\mu_i}$ and h satisfies FEI$_\eta$, with $\eta = \prod_{i=1}^{l} \eta_i$ and $\mathbb{E}\, \eta_i = \mathbb{E}_{\mu_i} g_i$, then f satisfies FEI$_\mu$.

Outline

Subcube Partitions

- Given $\alpha : [n] \to \{-1, +1, *\}$, *subcube* $C_\alpha := \{x \in \{0,1\}^n \ : \ \alpha(i) \neq * \implies x_i = \alpha(i)\}$.
- If $A := \{i \in [n] : \alpha(i) \neq *\}$, we also denote C_α by (A, α). The *co-dimension* of $C_\alpha = (A, \alpha)$ is $|A|$.
- A *subcube partition* $\mathcal{C} = \{C_1, \ldots, C_m\}$ of $\{0,1\}^n$ computes a function $f : \{0,1\}^n \to \{+1, -1\}$ if f is constant on each C_i.
- We denote by $L(f)$ the minimum number of subcubes in a subcube partition that computes f.
- Leaves of a decision tree computing f define a subcube partition that computes f.

Entropy from Concentration

Theorem

Let $f : \{0,1\}^n \to \{+1,-1\}$ depend on all its variables and be computed by a subcube partition $\mathcal{C}$ of size $L(f)$. Then, for some absolute constant $c > 1$,

$$\mathbb{H}(f) \leqslant c \cdot \log L(f).$$

It is well-known and easy to see that $\mathsf{Inf}(f) \leqslant \log L(f)$ for all f.

Proof idea:

- Show most Fourier mass is concentrated in a small set $\mathcal{B}$ of coefficients.
- Entropy within that set is bounded above by $\log \mathcal{B}$.
- If the leftover mass is small, say $< 1/n$, the leftover entropy is at most 1.

A (weak) concentration bound for subcube partitions

Lemma: Let $\mathcal{C} = \{(A_i, \alpha_i) : 1 \leqslant i \leqslant L\}$ compute f. Then $\forall t$, $\exists \mathcal{B}_t \subseteq 2^{[n]}$ such that (i) $|\mathcal{B}_t| \leqslant 2^{2t}$ and (ii) $\sum_{S \notin \mathcal{B}_t} \widehat{f}^2(S) \leqslant L \cdot 2^{-t}$.

Proof:

- $\mathcal{B}_t := \{S : \exists\, i \, |A_i| \leqslant t \text{ such that } S \subseteq A_i\}$
- Main point: only sets $A_I \supseteq S$ contribute to $\hat{f}(S)$
- $g \equiv \sum_{|A_i|>t} \beta_i \phi_i$: restriction of f to subcubes with $|A_i| > t$
- $\sum_{S \notin \mathcal{B}_t} \widehat{f}^2(S) = \sum_{S \notin \mathcal{B}_t} \widehat{g}^2(S) \leqslant \sum_{S} \widehat{g}^2(S) = 2^{-n} \sum_{|A_i|>t} |C_i| = \sum_{|A_i|>t} 2^{-|A_i|} < 2^{-t} L.$
- Since $\sum_i 2^{-|A_i|} = 1$, $|\{i : |A_i| \leqslant t\}| \leqslant 2^t$
- $|\mathcal{B}_t| \leqslant \sum_{|A_i| \leqslant t} 2^{|A_i|} \leqslant 2^t \cdot |\{i : |A_i| \leqslant t\}| \leqslant 2^{2t}$

Entropy upper bound on subcube partitions

- Fix $t := \log(Ln)$ in the lemma
- $$\begin{aligned} \mathbb{H}(f) &= \sum_S \hat{f}^2(S) \log \frac{1}{\hat{f}^2(S)} \\ &= (1 - 1/n)\, \mathbb{H}(\hat{f}^2(S) \;:\; S \in \mathcal{B}_t) + (1/n)\, \mathbb{H}(\hat{f}^2(S) \;:\; S \notin \mathcal{B}_t) + \mathrm{H}(1/n) \\ &\leqslant (1 - 1/n) \log |\mathcal{B}_t| + 1/n \cdot n + \mathrm{H}(1/n) \\ &\leqslant 2t + 1 + \mathrm{H}(1/n) \\ &\leqslant 2 \log L + 2 \log n + 2. \end{aligned}$$
- Lemma: Suppose $f : \{0,1\}^n \to \{+1, -1\}$ depends on all its variables. Then any subcube partition that computes f must have at least $n+1$ subcubes in it. That is $L \geqslant n + 1$.
- Thus, $\mathbb{H}(f) \leqslant 4 \log L + 2$.

Outline

FEI as a Coding Problem

Wan, Wright, Wu 2014 :

- Construct a *prefix-free code* c over alphabet Σ to minimize the *expected length* of a codeword under the distribution $\widehat{f}^2(S)$: $\mathbb{E}_{\hat{f}^2} |c(S)|$.
- Shannon's Source Coding Theorem: $\mathbb{H}(f) \log |\Sigma| \leqslant \mathbb{E}_{\hat{f}^2} |c(S)|$.
- Goal: construct such code with expected length $O(\mathsf{Inf}(f))$.
- By using the $\lceil \log n \rceil$-bit rep for each $i \in S$ and appending a terminating symbol, we get a prefix-free code with $|\Sigma| = 3$. The expected length of this code is $\lceil \log n \rceil \cdot \mathbb{E}_{\hat{f}^2} |S| + 1 = \lceil \log n \rceil \cdot \mathsf{Inf}(f) + 1$. This gives $\mathbb{H}(f) \leqslant \lceil \log n \rceil \cdot \mathsf{Inf}(f) + 1$.
- WWW 2014 give a protocol for encoding a S using a decision tree for f and prove that the expected length of the resulting prefix code is O(average depth of the DT). Note that $\mathsf{Inf}(f) \leqslant$ average depth of a DT computing f. This reproduces a result from [CKLS 2013].

Outline

Summary

- Bounds on Noise Stability of Derivatives – symmetric functions
- Composability properties of the FEI conjecture – read-once formulas
- Concentration Bounds – weaker forms of the conjecture using DT and subcube partition complexities instead of Influence
- Coding – weaker forms using DT complexity