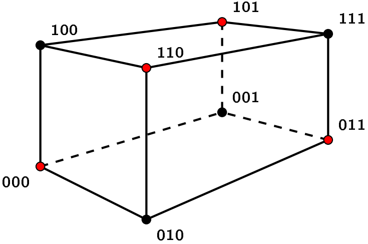
On Fourier Entropy-Influence Conjecture

Nitin Saurabh

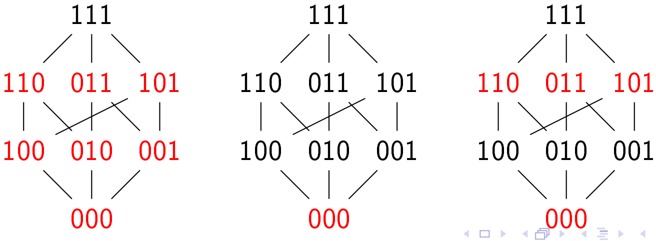
Technion - IIT, Haifa

April 10, 2021

Boolean Functions



- $f \; : \; \{0,1\}^n \to \{0,1\}, \quad 0 = \text{False}, \quad 1 = \text{True}$
- Examples: **AND**, **OR**, **Parity**



Polynomials

- unique (multilinear) polynomial rep. over $\mathbb{R}$

  Change of basis

- $0 \rightarrow 1, \;\; 1 \rightarrow -1, \qquad x \mapsto 1 - 2x$
- $f : \{+1, -1\}^n \rightarrow \{+1, -1\}$
- f – uniquely reprsentable by multilinear polynomials over $\mathbb{R}$

Examples

- **Parity**: $\{+1,-1\}^n \to \{+1,-1\}$

$$x_1 \cdot x_2 \cdots x_n.$$

- **AND**: $\{+1,-1\}^n \to \{+1,-1\}$

$$1 - 2\prod_{i=1}^{n}\left(\frac{1-x_i}{2}\right) = 1 - \frac{1}{2^{n-1}} + \sum_{S\subseteq[n]} \frac{(-1)^{|S|+1}}{2^{n-1}} \prod_{i\in S} x_i$$

- **OR**: $\{+1,-1\}^n \to \{+1,-1\}$

$$-1 + 2\prod_{i=1}^{n}\left(\frac{1+x_i}{2}\right) = -1 + \frac{1}{2^{n-1}} + \sum_{S\subseteq[n]} \frac{1}{2^{n-1}} \prod_{i\in S} x_i$$

Fourier Transform

- $\chi_S(x) := \prod_{i \in S} x_i, \quad S \subseteq [n]$
- Orthonormal w.r.t.
  $\langle f, g \rangle \ := \ \mathbb{E}_{x \sim \{1,-1\}^n}[f(x) \cdot g(x)] = \frac{1}{2^n} \sum_x f(x) \cdot g(x)$
  $\langle \chi_S, \chi_T \rangle = 1$ if $S = T$, otherwise $= 0$.
- *Unique* Fourier expansion,

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i,$$

where $\hat{f}(S) = \mathbb{E}_x[f \cdot \chi_S] = \frac{1}{2^n} \sum_x f(x) \prod_{i \in S} x_i$

Fourier Transform & Shannon entropy

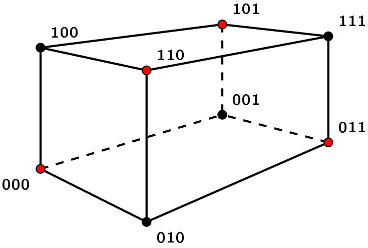
- [Parseval's Theorem] $\sum_{S\subseteq[n]} \hat{f}(S)^2 = \mathbb{E}_x[f(x)^2] = 1$
- Fourier distribution : $\left\{\hat{f}(S)^2\right\}_{S\subseteq[n]}$
- (Shannon) Entropy of f

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

$$\mathbb{H}(\mathbf{AND}) = \mathbb{H}(\mathbf{OR}) \leq \frac{4n+2}{2^{n-1}}$$

$$\mathbb{H}(\mathbf{Parity}) = 0$$

Influence / Average Sensitivity



- sensitivity of f at x: $s(f,x) := \#\left\{i \in [n] \mid f(x) \neq f(x^i)\right\}$
  x^i : obtained by flipping i-th bit in x
- $\mathrm{Inf}(f) =$ average sensitivity of $f := \mathbb{E}_x[s(f,x)]$
- $\mathrm{Inf}_i(f) := \Pr_x[f(x) \neq f(x^i)]; \quad \mathrm{Inf}(f) = \sum_{i=1}^n \mathrm{Inf}_i(f)$

  $\mathrm{Inf}_i(\mathbf{Parity}) = 1; \quad \mathrm{Inf}(\mathbf{Parity}) = n$

  $\mathrm{Inf}_i(\mathbf{AND}) = \frac{1}{2^{n-1}}; \quad \mathrm{Inf}(\mathbf{AND}) = \frac{n}{2^{n-1}}$

Fourier Entropy-Influence Conjecture

[Friedgut-Kalai, 1996]

$\exists\ C > 0$ s.t. for every $f : \{+1, -1\}^n \to \{+1, -1\}$,

$$\mathbb{H}(f) \leq C \cdot \mathrm{Inf}(f).$$

$$\sum_{S \subseteq [n]} \hat{f}(S)^2 \log \frac{1}{\hat{f}(S)^2} \leq C \cdot \sum_{S \subseteq [n]} |S| \hat{f}(S)^2.$$

$$\mathbb{H}(\mathbf{OR}) = \mathbb{H}(\mathbf{AND}) \leq \tfrac{4n+2}{2^{n-1}} \leq \tfrac{5n}{2^{n-1}} \leq 5 \cdot \mathrm{Inf}(\mathbf{AND}) = \mathrm{Inf}(\mathbf{OR})$$

$$\mathbb{H}(\mathbf{Parity}) = 0 \leq n = \mathrm{Inf}(\mathbf{Parity})$$

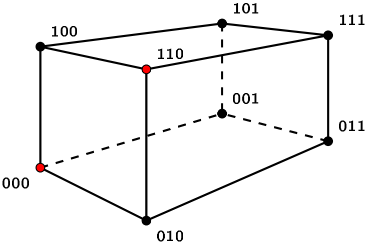
Current status of FEIC

- **Goal :** for all Boolean functions – seems to be *really* hard!
- **Approach 1:** for special classes of functions
  - Random DNFs [KLW'10], and Random LTFs [CKKLS'18]
  - Symmetric functions and read-once decision trees [OWZ'11]
  - Read-once Formulas [OT'14, CKLS'15]
  - Read-k Decision trees [WWW'14, Shalev'18]
- **Approach 2:** weak version of FEIC
  - Min-entropy-Influence conjecture for Read-k DNFs [ACKSdW'20]
  - (almost) Min-entropy-Influence conjecture [KKLMS'20]

Current status of FEIC (contd.)

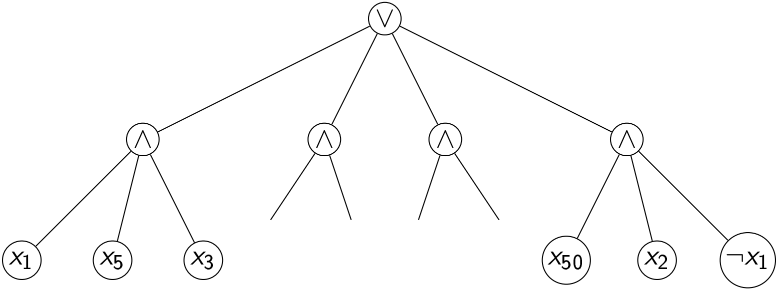
- **Approach 3:** weaker upper bounds on Fourier entropy
  - $\mathrm{Inf}(f) \cdot \log n$ and $\log L_1(f)$ [Folklore]
  - average depth – $\oplus$-decision tree [CKLS'15]
  - $\sqrt{n}$ for LTFs (degree-d PTFs: $O_d(n^{1-\frac{1}{4d+6}})$) [CKLS'15]
  - $\mathrm{Inf}(f) \cdot \log \mathrm{s}(f)$ [GSTW'16]
  - average unambiguous $\oplus$-certificate complexity [ACKSdW'20]

Certificate Complexity



- certificates = monochromatic subcubes = **AND** of literals
- $\mathsf{C}(f,x)$ – *certificate* complexity of f at x
  min. # bits in x that must be seen to certify function value
  $\mathsf{C}(f,000)=3$, $\mathsf{C}(f,011)=1$, $\mathsf{C}(\textbf{Parity},*)=n$
- $\mathsf{C}(f) := \max_x \mathsf{C}(f,x)$, $\mathsf{C}^1(f) := \max_{x\in f^{-1}(1)} \mathsf{C}(f,x)$,
  $\mathsf{C}^0(f) := \max_{x\in f^{-1}(0)} \mathsf{C}(f,x)$
- $\mathsf{s}(f,x) \leq \mathsf{C}(f,x)$

Certificates & DNFs

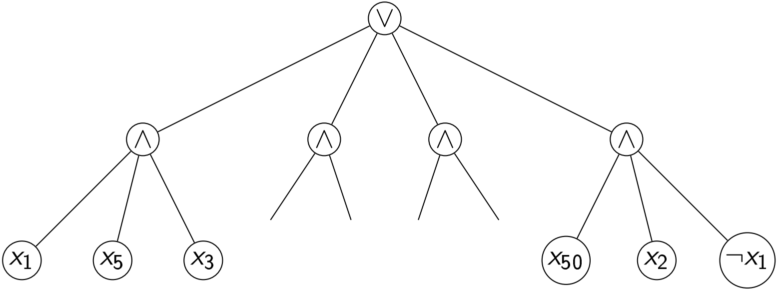


$$\text{t-term DNF} : f = T_1 \vee T_2 \vee \cdots \vee T_t$$

- Suppose each T_i sets at most k literals.
- Is $\mathbb{H}(f) = O(k)$? Is $\mathbb{H}(f) = O(\log t)$?
  Is $\mathbb{H}(f) \leq \min\left\{\mathrm{C}^1(f), \mathrm{C}^0(f)\right\}$?
- Both statements are weaker than FEIC
- *Both* imply Mansour's conjecture!

Mansour's conjecture for DNFs

- (weak) *Mansour's* Conjecture (1995)



$$\text{t-term DNF} : f = T_1 \vee T_2 \vee \cdots \vee T_t$$

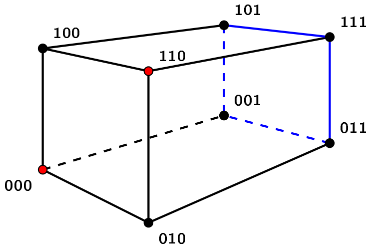
$$\text{Then, } \exists \ \ p : \{\pm 1\}^n \to \mathbb{R} \text{ with } t^{O(\log 1/\epsilon)} \text{ monomials}$$
$$\text{s.t. } \ \mathbb{E}_x[(f(x) - p(x))^2] \leq \epsilon.$$

- (weak) existence of p with $t^{O(1/\epsilon)}$ monomials

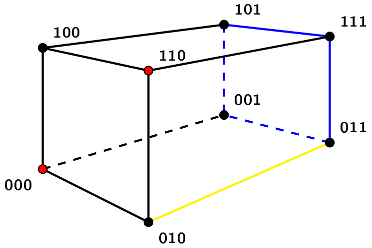
Unambiguous Certificates

- *Unambiguous* certificate $\mathcal{C} = \{C_1, \ldots C_t\}$
  collection of certificates – together partitions the space

Unambiguous Certificate

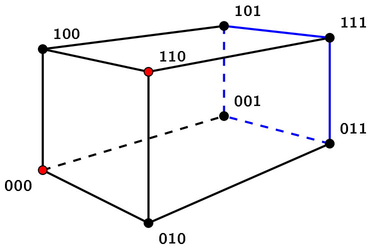


Unambiguous



Not unambiguous

Avg. Unambiguous Certificate Complexity



- unambiguous certificate $= \{000, 100, 110, 010, **1\}$

  avg. unambiguous certificate complexity
  $$= \frac{1}{2^3}(3+3+3+3+4*1) = 2$$

- avg. unambiguous cert. complexity := avg. # bits set by a certificate

Our Result

Theorem [Arunachalam-Chakraborty-Koucký-S-de Wolf'18]

For any $f : \{\pm 1\}^n \to \{\pm 1\}$,
$\mathbb{H}(f) \leq 2 \cdot$ avg. unambiguous certificate complexity.

Proof Outline

- $f \xrightarrow{\text{Tensorization}} f^M$ such that
  - $\mathbb{H}(f^M) = M \cdot \mathbb{H}(f)$
  - $\mathsf{aUC}(f^M) = M \cdot \mathsf{aUC}(f^M)$
- Find a "small" set $\mathcal{B}$ of heavy Fourier coefficients of f^M

  "small" $\sim 2^{\mathsf{aUC}(f^M)}$
- Fourier weight not in $\mathcal{B}$ is *negligible* (at most ε)
- Therefore, $\mathbb{H}(f^M) \leq \log |\mathcal{B}|$
- Steps (2) and (3) uses tools from Information Theory
  – Shannon's AEP Theorem

Tensorization

- For any $f : \{\pm 1\}^n \to \{\pm 1\}$ and $M \in \mathbb{N}$, define $f^M : \{\pm 1\}^{Mn} \to \{\pm 1\}$

  $$f^M(\tilde{x}^1, \ldots, \tilde{x}^M) := f(x_{11}, \ldots, x_{1n}) \cdot f(x_{21}, \ldots, x_{2n}) \cdot f(x_{M1}, \ldots, x_{Mn})$$

- Claim: $\mathbb{H}(f^M) = M \cdot \mathbb{H}(f)$
- Claim: $\mathrm{Inf}(f^M) = M \cdot \mathrm{Inf}(f)$
- tensorized version $\mathcal{C}^M$ of unambiguous certificate $\mathcal{C}$ of f := direct product of M independent copies of $\mathcal{C}$
- Claim: $\mathcal{C}^M$ unambiguous certificate of f^M

Asymptotic Equipartition Property

AEP Theorem (Shannon)

i.i.d $X_1, X_2, \ldots, X_M \sim X$, then

$$-\frac{1}{M}\log p(X_1, \ldots, X_M) \to \mathbb{H}(X)$$

in probability as $M \to \infty$.

1. $\epsilon > 0$. *typical set* $T_\epsilon^{(M)}$– set of sequences $(x_1, x_2, \ldots, x_M)$ s.t.

$$2^{-M(\mathbb{H}(X)+\epsilon)} \leq p(x_1, x_2, \ldots, x_M) \leq 2^{-M(\mathbb{H}(X)-\epsilon)}$$

2. $|T_\epsilon^{(M)}| \leq 2^{M(\mathbb{H}(X)+\epsilon)}$
3. $\Pr\{T_\epsilon^{(M)}\} > 1 - \epsilon$ for M sufficiently large.

Proof

- Consider tensorized version of f, $f^M : \{\pm 1\}^{Mn} \to \{\pm 1\}$ tensorized version of an unambigous cert. $\mathcal{C}$
- Define r.v. $\mathbf{C}$ supported on unambiguous certificate $\mathcal{C} = \{C_1, \ldots, C_t\}$ s.t. $\mathbf{C} = C_i$ with prob. equal fraction of points covered by C_i
  If $C_i =$ **AND** of k-literals, then prob. of choosing C_i equals 2^{-k}.
- typical set $T_\delta^{(M)}(\mathbf{C})$ for some $\delta > 0$.
- Proof idea: define a "small" set $\mathcal{B}$ of Fourier coefficients of f^M using the typical set $T_\delta^{(M)}(\mathbf{C})$ s.t. Fourier weight on this set is at least $1 - \delta$.

- Define $\mathcal{B}$ – M-tuple of subsets of $[n]$

$$\mathcal{B} := \{(S_1, \ldots, S_M) \mid S_i \subseteq \mathrm{vars}(C_i) \subseteq [n],\ (C_1, \ldots, C_M) \in T_\delta^{(M)}(\mathbf{C})\}.$$

- Claim 1: $|\mathcal{B}| \leq 2^{2M(\mathrm{aUC}(f,\mathcal{C})+\delta)}$.
- Claim 2: $\sum_{(S_1,S_2,\ldots,S_M)\notin\mathcal{B}} \widehat{f^M}(S_1 \cup \cdots \cup S_M)^2 \leq \delta$.

$$\begin{aligned}\mathbb{H}(f^M) &\leq \mathbb{H}(\widehat{f^M}(S)^2 : S \in \mathcal{B}) + \delta \cdot \mathbb{H}(\widehat{f^M}(S)^2 : S \notin \mathcal{B}) + \mathrm{H}(\delta) \\ &\leq \log|\mathcal{B}| + \delta \cdot (Mn) + \mathrm{H}(\delta) \\ &\leq 2M(\mathrm{aUC}(f,\mathcal{C}) + \delta) + \delta Mn + \mathrm{H}(\delta)\end{aligned}$$

Taking limit as $M \to \infty$,

$$\mathbb{H}(f) \leq 2 \cdot \mathrm{aUC}(f, \mathcal{C}).$$

Proof of Claim 1

- Claim: $|\mathcal{B}| \leq 2^{2M(\mathrm{aUC}(f,\mathcal{C})+\delta)}$.
- By AEP property (2), # typical sequences $\leq 2^{M(\mathbb{H}(\mathbf{C})+\delta)}$
- Each contribute exactly $2^{\sum_i \mathrm{vars}(C_i)}$ tuples
- However, $2^{\sum_i \mathrm{vars}(C_i)} = \Pr[\mathbf{C}_1 = C_1, \ldots \mathbf{C}_M = C_M]^{-1}$
- Since $(C_1, \ldots, C_M) \in T_\delta^{(M)}(\mathbf{C})$, by AEP property (1),
  $\Pr[\mathbf{C}_1 = C_1, \ldots \mathbf{C}_M = C_M]^{-1} \leq 2^{M(\mathbb{H}(\mathbf{C})+\delta)}$
- $|\mathcal{B}| \leq 2^{2M(\mathbb{H}(\mathbf{C})+\delta)}$
- $\mathbb{H}(\mathbf{C}) = \mathrm{aUC}(f,\mathcal{C})$

Proof of Claim 2

- Claim: $\sum_{(S_1,S_2,\ldots,S_M)\notin\mathcal{B}} \widehat{f^M}(S_1 \cup \cdots \cup S_M)^2 \leq \delta$
- Recall, $\mathcal{C}^M$ unambiguous certificate for f^M

  $\rho \in \mathcal{C}^M$ be a certificate of f^M

  $\mathbb{1}_\rho(z)$ – $\{0,1\}$-indicator variable
  $= 1$ iff z is consistent with the certificate ρ

$$f^M(z) = \sum_{\rho\in\mathcal{C}^M} f^M(\rho)\cdot\mathbb{1}_\rho(z)$$
$$= \sum_{\rho\in T_\delta^{(M)}(\mathbf{C})} f^M(\rho)\cdot\mathbb{1}_\rho(z) + \underbrace{\sum_{\rho\notin T_\delta^{(M)}(\mathbf{C})} f^M(\rho)\cdot\mathbb{1}_\rho(z)}_{:=g(z)}.$$

Obs: $(S_1, \ldots, S_M) \notin \mathcal{B} \implies \widehat{f^M}(S_1 \cup \cdots \cup S_M)$ gets contribution only from $\rho \notin T_\delta^{(M)}(\mathbf{C})$

- $\sum_{(S_1, S_2, \ldots, S_M) \notin \mathcal{B}} \widehat{f^M}(S_1 \cup \cdots \cup S_M)^2 \leq \sum_T \hat{g}(T)^2$
- From Parseval's, $\sum_T \hat{g}(T)^2 = \mathbb{E}_z[g(z)^2]$
- However, $g : \{\pm 1\}^{Mn} \to \{\pm 1, 0\}$
- Therefore, $\mathbb{E}_z[g(z)^2] = \Pr_z[z \notin T_\delta^{(M)}(\mathbf{C})]$
- By AEP property (3), $\Pr_z[z \notin T_\delta^{(M)}(\mathbf{C})] \leq \delta$

Open Problems

- Prove Mansour's Conjecture!
  - entropy $\leq$ bottom fan-in for DNFs ?
  - entropy $\leq$ log (top fan-in) ?
- FEIC for Linear Threshold Functions?
- Min-entropy-Influence conjecture for DNFs ?
- Prove our results bypassing AEP.

Towards Mansour's Conjecture

- $f = g_1 + \cdots + g_t$ s.t. g_i's are $\{0,1\}$ or $\{0,-1\}$ valued functions, and on any input *exactly* one of them evaluates to *non-zero*
- $\sum_i \Pr[g_i \neq 0] = 1$
- $\Pr[g_i \neq 0] = \sum_{S \subseteq [n]} \widehat{g_i}(S)^2$
- Is the following true?

$$\mathbb{H}(f) = O\left(\sum_{i=1}^{t} \sum_{S \subseteq [n]} \widehat{g_i}(S)^2 \log \frac{1}{\widehat{g_i}(S)^2} + \sum_{i=1}^{t} \Pr[g_i \neq 0] \log \frac{1}{\Pr[g_i \neq 0]} \right)$$

Thank You!

Questions?