The Sensitivity Conjecture and Fourier Analysis of Boolean functions

A brief walk through theoretical computer science

Debmalya Bandyopadhyay

Department of Mathematics and Statistics
Indian Institute of Science Education and Research, Kolkata

April 8, 2022

Summary of talk

1. Combinatorial Complexity
2. The Sensitivity Conjecture
3. Fourier Analysis of Boolean functions
4. Flat, Homogeneous, Boolean

Combinatorial Complexity

Boolean Functions

Definition

A Boolean function is a function $f : \{0,1\}^n \to \{0,1\}$ or $f : \{\pm 1\}^n \to \{\pm 1\}$.

- It maps a string of n Boolean variables to a single Boolean value.

Use of Boolean functions

Boolean functions are widely used in different contexts, such as:

- In circuit designing, it can represent how a circuit behaves based on n inputs and one output.
- In graph theory one can identify a graph G with $|V(G)|$ vertices with a $\binom{|V(G)|}{2}$ vector that indicates which edges are present. Then the Boolean function can be an indicator function for a particular property of a graph, such as $f(G) = 1$ iff G is connected.
- In social choice theory, a Boolean function can be identified with a "voting rule" for an election with two candidates 0 and 1

Examples of Boolean functions

- $AND_n(x) = 1$ iff $|x| = n$, $OR_n(x) = 1$ iff $|x| \geq 1$
- Rubinstein's function: $f : \{0,1\}^{k^2} \to \{0,1\}$
  $f(x) = 1 \iff \exists$ one block B of k variables in x s.t.
  $B = 0^l 1 1 0^{k-(l+2)}$
- For a graph G, with $|V(G)| = n$, let $f : \{0,1\}^{\binom{n}{2}} \to \{0,1\}$ s.t.
  $f(x) = 1 \iff G$ is connected.

Combinatorial Complexity

The goal of combinatorial complexity of a Boolean function:

- Understand the "complexity" of a function
- There can be various measures of "complexity" of a function
- One also want to understand
  - How the measures behave for different classes of Boolean functions
  - How is the relation between various measures.

Combinatoiral Measure: Sensitivity

Definition

Given a Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ the **sensitivity** of f at a point x, $s(f, x) = \#\{i : f(x) \neq f(x^i)\}$

$$\text{Sensitivity of } f,\ s(f) := max_x s(f, x)$$

Example

Let $f : \mathbb{F}_2^8 \to \mathbb{F}_2$, such that $f(x) = 1 \iff |x| = 4$ or 5
Take $x = 11100000$, then $f(x) \neq f(x^i)$ for $4 \leq i \leq 8$

$$s(f, x) = 5 = s(f)$$

Combinatorial Measure: Block Sensitivity

Definition

Given a Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ the **block sensitivity** of f at a point x, is the maximum number b such that there are disjoint sets $B_1, B_2, ..., B_b$ for which $f(x) \neq f(x^{B_i})$

$$bs(f) := max_x bs(f, x)$$

Example

Let $f : \mathbb{F}_2^8 \to \mathbb{F}_2$, such that $f(x) = 1 \iff |x| = 4$ or 5
Take $x = 11110000$, then Sensitive blocks -
$\{1\}, \{2\}, \{3\}, \{4\}, \{5, 6\}, \{7, 8\}$

$$bs(f, x) = 6 = bs(f)$$

Combinatorial Measure: Degree of a Boolean function

Theorem (Fourier Expansion Theorem)

Every $f : \{\pm 1\}^n \to \mathbb{R}$ *can be represented uniquely as a multilinear polynomial*

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

where $\chi_S(x) = \prod_{i \in S} x_i$

Definition

$\deg(f) := \max\{|S| : \hat{f}(S) \neq 0\}$

Lemma (Nisan-Szegedy)

For any Boolean function f, $bs(f) \leq 2deg(f)^2$

Other Combinatorial Measures ...

There are a lot of other combinatorial complexity measures like Entropy, Influence, Decision tree complexity, Certificate Complexity, etc.

Challenging questions in this area:

How are the various measures related to each other for various classes of Boolean functions?

For example,

(Upper bound) Is $bs(f) \leq O(s(f)^2)$?
(Lower Bound) Is there an f such that $bs(f) \geq \Omega(s(f)^2)$?

Already established relations between complexity measures

Table 1: Best known separations between complexity measures

	D	R_0	R	C	RC	bs	s	λ	Q_E	deg	Q	$\widetilde{\deg}$
D		2, 2 [ABB+17]	2, 3 [ABB+17]	2, 2 $\wedge \circ \vee$	2, 3 $\wedge \circ \vee$	2, 3 $\wedge \circ \vee$	3, 6 [BHT17]	4, 6 [ABB+17]	2, 3 [ABB+17]	2, 3 [GPW18]	4, 4 [ABB+17]	4, 6 [ABB+17]
R_0	1, 1 $\oplus$		2, 2 [ABB+17]	2, 2 $\wedge \circ \vee$	2, 3 $\wedge \circ \vee$	2, 3 $\wedge \circ \vee$	3, 6 [BHT17]	3, 6 [BHT17]	2, 3 [ABB+17]	2, 3 [GJPW18]	3, 4 [ABB+17]	4, 6 [ABB+17]
R	1, 1 $\oplus$	1, 1 $\oplus$		2, 2 $\wedge \circ \vee$	2, 3 $\wedge \circ \vee$	2, 3 $\wedge \circ \vee$	3, 6 [BHT17]	3, 6 [BHT17]	$\frac{3}{2}$, 3 [ABB+17]	2, 3 [GJPW18]	$\frac{8}{3}$, 4 [Tal19]	4, 6 [ABB+17]
C	1, 1 $\oplus$	1, 1 $\oplus$	1, 2 $\oplus$		2, 2 [GSS13]	2, 2 [GSS13]	2.22, 5 [BHT17]	2.22, 6 [BHT17]	1.15, 3 [Amb13]	1.63, 3 [NW95]	2, 4 $\wedge$	2, 4 $\wedge$
RC	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$		$\frac{3}{2}$, 2 [GSS13]	2, 4 [Rub95]	2, 4 $\wedge$	1.15, 2 [Amb13]	1.63, 2 [NW95]	2, 2 $\wedge$	2, 2 $\wedge$
bs	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$		2, 4 [Rub95]	2, 4 $\wedge$	1.15, 2 [Amb13]	1.63, 2 [NW95]	2, 2 $\wedge$	2, 2 $\wedge$
s	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$		2, 2 $\wedge$	1.15, 2 [Amb13]	1.63, 2 [NW95]	2, 2 $\wedge$	2, 2 $\wedge$
λ	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$		1, 1 $\oplus$	1, 2 $\oplus$	1, 1 $\oplus$	1, 2 $\oplus$
Q_E	1, 1 $\oplus$	1.33, 2 $\bar{\wedge}$-tree	1.33, 3 $\bar{\wedge}$-tree	2, 2 $\wedge \circ \vee$	2, 3 $\wedge \circ \vee$	2, 3 $\wedge \circ \vee$	3, 6 [BHT17]	3, 6 [BHT17]		2, 3 [ABK16]	2, 4 $\wedge$	4, 6 [ABK16]
deg	1, 1 $\oplus$	1.33, 2 $\bar{\wedge}$-tree	1.33, 2 $\bar{\wedge}$-tree	2, 2 $\wedge \circ \vee$	2, 2 $\wedge \circ \vee$	2, 2 $\wedge \circ \vee$	2, 2 $\wedge \circ \vee$	2, 2 $\wedge$	1, 1 $\oplus$		2, 2 $\wedge$	2, 4 $\wedge$
Q	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	2, 2 [ABK16]	2, 3 [ABK16]	2, 3 [ABK16]	3, 6 [BHT17]	3, 6 [BHT17]	1, 1 $\oplus$	2, 3 [ABK16]		4, 6 [ABK16]
$\widetilde{\deg}$	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	2, 2 [BT17]	2, 2 [BT17]	2, 2 [BT17]	2, 2 [BT17]	2, 2 [BT17]	1, 1 $\oplus$	1, 1 $\oplus$	1, 1 $\oplus$	

The Sensitivity Conjecture

Sensitivity Conjecture

Theorem (Sensitivity theorem (Hao 2019):)

$\exists C > 0$ *such that* $bs(f) \leq s(f)^C$ *for every Boolean function* f.

Proof of Sensitivity Conjecture

- Existing techniques were either combinatorial or analytical
- Result known for special classes of Boolean functions: Symmetric, Graph Properties, Minterm-transitive
- Gotsman-Linial gave a path towards proving the conjecture via graph theory
- Hao Huang (2019) proved the conjecture using the Gotsman-Linial Technique

Theorem (Hao 2019)

For any Boolean function, deg$(f) \leq s(f)^2$

Gotsman-Linial (GL92) Observation

Proving an effective upper bound for $deg(f)$ in terms of $s(f)$ is equivalent to:

- Let Q_n be a graph on 2^n vertices indexed by $\{\pm 1\}^n$ having an edge between two vertices x, y iff $\#\{i : x_i \neq y_i\} = 1$ and suppose $\Delta(G)$ is the maximum degree of a graph G. For an induced subgraph G of Q_n with strictly greater than half the vertices (i.e. 2^{n-1} vertices), find a lower bound of $\Delta(G)$ in terms of n.

Hao proved

Lemma

Let G be a $2^{n-1}+1$ vertex induced subgraph of Q_n. Then $\Delta(G) \geq \sqrt{n}$.

- So combining with GL92 technique we have the proof of sensitivity conjecture.

Proof of Hao's Lemma

Theorem (Cauchy's Interlace theorem:)

Let A be a symmetric $n \times n$ matrix and B be an $m \times m$ principal submatrix. If eigenvalues of A are $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ and eigenvalues of B are $\mu_1 \geq \mu_2 \geq ...\mu_m$, then for $1 \leq i \leq m$,

$$\lambda_i \geq \mu_i \geq \lambda_{i+n-m}$$

- Hao iteratively defines symmetric matrices A_n where A_n is a $2^n \times 2^n$ matrix with eigenvalues $\sqrt{n}$ with multiplicity 2^{n-1} and $-\sqrt{n}$ with multiplicity 2^{n-1} .
- Entries of A_n are 0 (whenever the adjacency matrix of Q_n) has no edge, and 1 or -1 otherwise.

Proof Continued

- Suppose H is an m-vertex undirected graph, and A is a symmetric matrix with entries in $\{0, \pm 1\}$ and whose rows and columns are indexed by $V(H)$, and whenever u and v are non-adjacent in H, $A_{u,v} = 0$. Then $\Delta(H) \geq \lambda_1(A)$
- He takes H as a $2^{n-1} + 1$ vertex induced subgraph of Q_n and the principal submatrix A_H of A_n naturally induced by H to apply the above result.
- By Cauchy Interlace theorem, $\lambda_1(A_H) \geq \lambda_{2^{n-1}}(A_n) = \sqrt{n}$

Further Developments

There have been some reworkings of Huang's proof in the last two years.

- Knuth (2019) proved a computational improvement of Huang's proof.
- Laplante-Naserasr-Sunny (2020) gave a purely linear algebraic construction that improved the result to $deg(f) \leq s_0(f)s_1(f)$ which gives us $bs(f) \leq s_0(f)^2 s_1(f)^2$.
- The best possible separation between sensitivity and block sensitivity we have so far is witnessed by Rubinstein's function (1995), for which $bs(f) \in O(s(f)^2)$. So,

Open Problem:

For every Boolean function f, $bs(f) \leq s(f)^2$

Fourier Analysis of Boolean functions

Fourier expansion

The Fourier expansion of a function $f : \{\pm 1\}^n \to \mathbb{R}$ is representation of the function as a real, multilinear polynomial.

For example,

the maximum function on 2 bits:

$$max_2(x_1, x_2) = \begin{cases} -1 & x_1 = x_2 = -1 \\ 1 & \text{otherwise} \end{cases}$$

Can be represented as $max_2(x_1, x_2) = \frac{1}{2} + \frac{1}{2}(x_1 + x_2) - \frac{1}{2}x_1x_2$ and this is the Fourier expansion of max_2.

Given an arbitrary Boolean function, we can uniquely find its multinomial representation in the following way.

The Fourier Expansion theorem

Theorem (Fourier Expansion Theorem)

Every $f : \{\pm 1\}^n \to \mathbb{R}$ *can be represented uniquely as a multilinear polynomial*

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) x^S$$

We often denote x^S *as* $\chi_S(x)$ *and* $deg(f) = max\{|S| : \hat{f}(S) \neq 0\}$

- We can add 2 such functions pointwise and also scalar multiply
- Set of all $f : \{\pm 1\}^n \to \mathbb{R}$ forms a vector space V of dimension 2^n
- Every function is a linear combination of the 2^n parity functions $(\chi_S(x) = \prod_{i \in S} x_i)$
- So they not only span V but also form a basis

Inner Products

We now introduce an inner product for pairs of function $f, g : \{\pm 1\}^n \to \mathbb{R}$

Inner product

$\langle f, g \rangle = 2^{-n} \sum_{x \in \{\pm 1\}^n} f(x)g(x) = \mathbb{E}_{x \in \{\pm 1\}^n}[f(x)g(x)]$

We now state two facts:

Fact 1: $\chi_S \chi_T = \chi_{S \Delta T}$ where $S \Delta T$ is the symmetric difference.
Fact 2:

$$\mathbb{E}[\chi_S(x)] = \mathbb{E}[\prod_{i \in S} x_i] = \begin{cases} 1 & S = \phi \\ 0 & \text{otherwise} \end{cases}$$

Orthonormal basis

From these 2 facts it follows:

Theorem

The 2^n parity functions $\chi_S(x) : \{\pm 1\}^n \to \{\pm 1\}$ form an orthonormal basis of V. So

$$\langle \chi_S, \chi_T \rangle = \begin{cases} 1 & S = T \\ 0 & \textit{otherwise} \end{cases}$$

Fourier Coefficients

Clearly $\langle f, \chi_S \rangle = \hat{f}(S)$

- When the range is $\mathbb{R}$ there is a 1-1 correspondence between a set of Fourier coefficients and some $f : \{\pm 1\}^n \to \mathbb{R}$
- However if the range is restricted to $\{\pm 1\}$ there **does not exist** a sufficient condition to check if a given set of constants correspond to a Boolean function

Parseval's and Plancherel's theorem

Theorem (Parseval's theorem)

For any function $f : \{\pm 1\}^n \to \mathbb{R}$*, we have*

$$\langle f, f\rangle = \mathbb{E}_{x\in\{\pm 1\}^n}[f(x)^2] = \sum_{S\subseteq[n]} \hat{f}(S)^2$$

Particularly, if the range of f is $\{\pm 1\}$ then we have $\sum_{S\subseteq[n]} \hat{f}(S)^2 = 1$

Theorem (Plancherel's theorem)

For any $f, g : \{\pm 1\}^n \to \mathbb{R}$*, we have*

$$\langle f, g\rangle = \mathbb{E}_x[f(x)g(x)] = \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S)$$

Entropy

- In information theory, the **entropy** of a random variable is the average level of "surprise" or "uncertainty" inherent in the variable's possible outcomes

Example:

Consider a biased coin Ber(p). The maximum surprise is for $p = 1/2$, when both outcomes are equally likely. In this case a coin flip has an entropy of one bit. The minimum surprise is for $p = 0$ or $p = 1$, when the event is known and the entropy is zero bits.

Shannon Entropy

Definition

suppose we have a discrete random variable X with possible outcomes $x_1, x_2, ...x_n$ which occur with probabilities $\{P(x_i)\}_{i=1}^{n}$. Then the Entropy of X is defined as

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i)$$

where the base of log is 2 to get the answer in units of bits (or shannons).

Fourier Entropy

- Parseval: $\sum_{S\subseteq[n]} \hat{f}(S)^2 = 1$
- the squared Fourier coefficients can be used as a probability distribution, over subsets $S \subseteq [n]$, and this is termed as Fourier distribution.

Definition (Shannon Entropy)

The Fourier entropy of f, denoted as $H(\hat{f}^2)$ is defined as the Shannon entropy of the Fourier Distribution.

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

Combinatorial Measure: Influence

Definition (Influence)

The total influence of a Boolean function is defined to be the expected size of a subset $S \subseteq [n]$ with the Fourier distribution.

$$Inf(f) = \sum_{S \subseteq [n]} |S| \hat{f}(S)^2$$

Combinatorially, the total influence is just the average sensitivity of f. For any $i \in [n]$ we can define the $Inf_i(f)$ as the probability that flipping the i-th bit on a random input flips the value of f. Then

$$Inf(f) = \sum_i Inf_i(f)$$

The FEI conjecture

- Fourier entropy is a measure of how spread out the Fourier distribution is over the 2^n coefficients.
- The total influence gives the idea how concentrated the Fourier distribution ison the higher degree terms.
- The FEI conjecture connects these two, informally saying that Boolean functions whose distribution is well spread out must have significant Fourier weight on the high degree monomials.

Fourier Entropy Influence Conjecture

$\exists C > 0$ such that for any Boolean function, $H(\hat{f}^2) \leq C \, \mathrm{Inf}(f)$

Implications of FEI Conjecture

Definition

A multilinear polynomial is called **flat** if all its nonzero coefficients have the same magnitude. It is called homogeneous when all the monomials bear the same degree.

FEI Conjecture implies:

- for a fixed constant $\epsilon \in (0, 1/2)$, a flat Boolean polynomial with degree d and sparsity $2^{\omega(d)}$ cannot ϵ-approximate any Boolean function.
- no homogeneous flat Boolean polynomial with degree d and sparsity $2^{\omega(d \log d)}$ can $1/3$-approximate a Boolean function.

Flat, Homogeneous, Boolean

Flat and homogeneous Boolean polynomials

Open question:

What are all the flat and homogeneous Boolean polynomials?

That is, to classify all polynomials $f : \{\pm 1\}^n \to \{\pm 1\}$ such that all the monomials have the same degree and all the coefficients have the same absolute magnitude.

Observation 1

Let's write $F = Cf$ where C is the positive constant, and f is a multilinear polynomial with coefficients ± 1, and its range is some nonzero integer $\pm k$. Therefore it forces $C = \frac{1}{k}$.

$\#$ positive terms $\neq \#$ negative terms

Notations and variable flips

- Monomial degree $= m$
- Total number of terms $= N$
- Total positive terms $= y$, Negative terms $= z$
- For any x_i, total positive terms $= y_i$, negative terms $= z_i$

* $y - z = k$
* $y_i - z_i = 0$ or k

A tiny inequality

Since each term is counted m times by the m variables present in it, we have $\sum_{i=1}^{n} \frac{y_i}{m} = y$ and $\sum_{i=1}^{n} \frac{z_i}{m} = z$

- how many variables should have unequal number of y_i and z_i?
- Is it possible that all the variables are balanced and have equal number of positive terms and negative terms holding them?

$$k = |y - z| = |\sum_{i=1}^{n} \frac{y_i - z_i}{m}| \leq \frac{1}{m} \sum_{i=1}^{n} |y_i - z_i|$$

What the inequality tells us is that at least m many variables have $|y_i - z_i| = k$.

Observation

- Parseval & Fourier distribution $\implies N = k^2$
- $y + z = k^2$ and $y - z = k$
- $y = \frac{k(k+1)}{2}$ and $z = \frac{k(k-1)}{2}$
- Thus, once we have k fixed, both y and z are fixed too!
- What values of k are permissible?

Fixing the coefficient and sparsity

Lemma:

Suppose that $f : \{\pm 1\}^n \to \{\pm 1\}$ has degree m. Then each $\hat{f}(S)$ is an integer multiple of 2^{1-m}.

- $\frac{1}{k}$ is an integer multiple of 2^{1-m}
- This forces k to be a power of 2
- $N = k^2$ is a power of 4.

The first non trivial such function

$n = 2m$

- Partition the m variables of the negative term into two nonempty parts, one that contains m_1 variables and the other with the remaining m_2 variables.
- Put the m_1 variables in the first positive term and the m_2 variables in the second positive term.
- Use $m - m_1$ new variables to complete the first positive term, and $m - m_2$ new variables to complete the second.
- We introduced $m_1 + m_2 = m$ new variables in the last step. These same m variables make the third positive term.

Examples:

$\frac{1}{2}(x_1x_3 - x_1x_2 + x_2x_4 + x_3x_4)$
$\frac{1}{2}(x_1x_4x_5 - x_1x_2x_3 + x_2x_3x_6 + x_4x_5x_6)$

Composition of Boolean functions

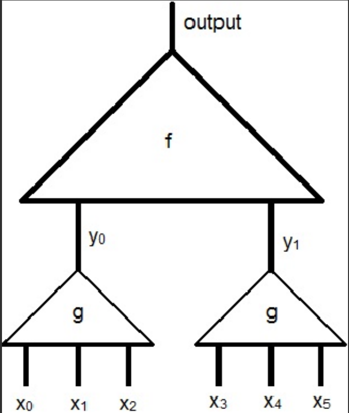

Figure: Composing a 3-variable Boolean fn with a 2-variable Boolean fn

Product and Composition

- Product of two flat/homogeneous Boolean functions($(m_1, k_1)(m_2, k_2)$) with disjoint input sets is flat/homogeneous $(m_1 + m_2, k_1 k_2)$ and Boolean.
- Composition of two flat/homogeneous Boolean functions is again flat, homogeneous and Boolean since it is merely sum of some products.

An observation

Important Observation

Product and composition of two FHB functions of the form $\frac{f_1(f_2+f_3)+f_4(f_2-f_3)}{2}$, where f_1, f_2, f_3, f_4 are flat and Boolean, can be written in the form $\frac{x_1(x_2+x_3)+x_4(x_2-x_3)}{2}$ via a suitable map such that each x_i are flat homogeneous and Boolean. (This is true for any number of variables)

Product reduction

It is easy to see for composition. For product, consider $f = \frac{f_1(f_2+f_3)+f_4(f_2-f_3)}{2}$ and $g = \frac{g_1(g_2+g_3)+g_4(g_2-g_3)}{2}$. Compute $f \times g$. Now consider the following map:

$x_1 = f_1$

$$x_2 = \frac{f_2[g_1(g_2+g_3)+g_4(g_2-g_3)]}{2}$$

$$x_3 = \frac{f_3[g_1(g_2+g_3)+g_4(g_2-g_3)]}{2}$$

$x_4 = f_4$

Then $f \times g = \frac{x_1(x_2+x_3)+x_4(x_2-x_3)}{2}$ via this map, and clearly all the x_i's are flat homogeneous and Boolean.

Conjecture

Conjecture:

Can any flat and homogeneous Boolean function be reduced to the 4-term function $\frac{x_1(x_2+x_3)+x_4(x_2-x_3)}{2}$, via variable transformations which are flat homogeneous and Boolean?

Starting with an arbitrary such function on 4^n terms, if we can reduce it to 4^{n-1} terms via such transformations, it is enough to conclude the conjecture.

Some further questions

- What is the relation between degree (m), number of variables (n) and Fourier sparsity (N)?
- Nisan-Szegedy (1992) $\implies m \geq \log_2 n - O(\log \log n)$
- Spectral norms $\implies n = \Omega(\log(N))$

Same Influence?

Can we probabilistically prove any two variables have the same influence?

$$Inf[f] = \sum_{|S| \subseteq [n]} |S| \hat{f}(S)^2$$

The End

Thanks for staying awake!