

On the Fourier spectrum of symmetric Boolean functions

Amir Shpilka

Technion and MSR NE

Based on joint work with

Avishay Tal

Theme: Analysis of Boolean functions

- Pick favorite representation: **Fourier Transform**
- Study structural properties of representation
 - (Anti-) Concentration of Fourier spectrum
 - Noise sensitivity
 - Degree
- **Too general?**
Study specific families of functions
- **This talk:**
Fourier spectrum of symmetric functions

Motivation

- Basic and natural question
- Similar questions have many applications:
 - Cryptography
 - Learning theory
 - Circuit lower bounds
 - Pseudorandomness
 - Voting theory
 - ...

Talk outline

- Definition of the problem
- Application: learning symmetric juntas
- Proof

Symmetric Boolean functions

- $g: \{0,1\}^k \rightarrow \{0,1\}$
- \forall permutation σ , $g(x_1, \dots, x_k) = g(x_{\sigma(1)}, \dots, x_{\sigma(k)})$
- $g(x)$ depends only on weight of x
- Equivalently:
- $g_{\mathbb{N}}: \{0, \dots, k\} \rightarrow \{0, 1\}$
- $g_{\mathbb{N}}(|x|) = g(x)$
- **Examples:** Parity, Majority, AND, OR, ...

Fourier transform of Boolean functions

- $g:\{0,1\}^k \rightarrow \{0,1\} \rightsquigarrow g:\{-1,1\}^k \rightarrow \{-1,1\}$
- $1 \rightarrow -1$ and $0 \rightarrow 1$
- **Fourier transform**: represent g as an n -variate polynomial over \mathbb{R} :

$$g(x_1, \dots, x_k) = \sum_{S \subseteq [k]} \hat{g}(S) \cdot \prod_{i \in S} x_i \triangleq \sum_{S \subseteq [k]} \hat{g}(S) \cdot x^S$$

- g symmetric $\Rightarrow \hat{g}(S)$ depends only on $|S|$
- **Example**:
 - Parity ^{k} = $x^{[k]} = x_1 \cdot x_2 \cdot \dots \cdot x_k$
 - MAJ(x_1, x_2, x_3) = $\frac{1}{2}(x_1 + x_2 + x_3 - x_1 \cdot x_2 \cdot x_3)$

Main questions

- $g(x_1, \dots, x_n) = \sum_{S \subseteq [k]} \hat{g}(S) \cdot \prod_{i \in S} x_i = \sum_{S \subseteq [k]} \hat{g}(S) \cdot x^S$
- g symmetric $\Rightarrow \hat{g}(S)$ depends only on $|S|$

- **Questions:** what can we say about

- Degree of Fourier representation of g

- Degree of minimal term in Fourier spectrum of g (if g is not Parity, \neg Parity).

In other words:

what is the minimal $S \neq \emptyset$ such that $\hat{g}(S) \neq 0$?

- Questions related as if $\hat{g}(\emptyset) = 0$ then
degree of minimal term in Four. spectrum of g
= degree of $g \oplus$ Parity

Correlation Immune functions

- Def: g is t -correlation immune if every subset of $\leq t$ variables is statistically independent of the value of g
- In other words: $\Pr_x[g(x)=1]$ is the same no matter how we fix any $\leq t$ variables
- Motivated by security of combining function for linear-feedback-shift-registers, e.g. when used to generate keys [Siegenthaler]
- Thm [Xiao Massey]: g is t -correlation immune iff $\hat{g}(S) = 0$ for all $0 \neq |S| \leq t$
- Recall: “What is the minimal $S \neq \emptyset$ such that $\hat{g}(S) \neq 0$? (if g is not Parity, \neg Parity)”

Learning juntas

(learning in the presence of irrelevant features)

- Unknown $h(x_1, \dots, x_n)$ that depends only on k variables (**k-junta**): $h(x_1, \dots, x_n) = g(x_{i_1}, \dots, x_{i_k})$
- Given uniformly random labeled examples $\langle x, h(x) \rangle$ find (w.h.p.) relevant variables
- **Goal**: do better than $n^k \cdot \text{poly}(n, 2^k)$
- **Lower bound**: $\geq 2^k$ samples necessary
- **[A. Blum]** most important problem in computational learning theory!

Learning juntas

(learning in the presence of irrelevant features)

- Given uniformly random labeled examples $\langle \mathbf{x}, h(\mathbf{x}) \rangle$, where $h(x_1, \dots, x_n) = g(x_{i_1}, \dots, x_{i_k})$, find (w.h.p.) relevant variables
- **Simple learning algorithm**: estimate $\hat{h}(S)$ for all $S \subseteq [n]$. If $\hat{h}(S) \neq 0$ then S is relevant!
- Estimating $\hat{h}(S)$ requires only $\text{poly}(n, k)$ samples (accuracy $\pm \exp(-k)$)
- **Wishful thinking**: If we only knew there is a small S with $\hat{g}(S) \neq 0$

Known results on Fourier spectrum of symmetric functions

- Degree of minimal term in Four. spectrum of g
 - [Mossell O'Donnell Servedio]: $|S| \leq 2k/3$
 - [Kolountzakis Lipton Markakis Mehta Vishnoi]: $|S| \leq k/\log(k)$
- Degree of Fourier representation of g
 - [von zur Gathen Roche]: $\deg(g) \geq k - k^{0.525}$
 - Conjecture [v-z-Gathen Roche]: $\deg(g) \geq k - O(1)$

Application: learning symmetric juntas

- **Goal:** do better than $n^k \cdot \text{poly}(n, 2^k)$
 - [MOS] general functions: $n^{\omega k / (\omega + 1)} \cdot \text{poly}(n, 2^k)$
if $\omega = 2$: $n^{\frac{2}{3}k} \cdot \text{poly}(n, 2^k)$
- Special case of **symmetric** junta:
 - [MOS]: $n^{\frac{2}{3}k} \cdot \text{poly}(n, 2^k)$
 - [KLMMV]: $n^{k / \log(k)} \cdot \text{poly}(n, 2^k)$
- Same algorithm, different analysis:
learn all Fourier coefficient $\hat{h}(S)$ until a nonzero coefficient is found

Our results

- Degree of minimal term in Fourier spectrum of g
 - $|S| \leq k^{0.525}$ (improves $|S| \leq k/\log(k)$)
 - Assuming ERH: $|S| \leq k^{0.5}$
- Learning symmetric juntas
 - Fourier learning algorithm runs in time $n^{k^{0.525}} \cdot \text{poly}(n, 2^k)$
(previous analysis: $n^{k/\log(k)} \cdot \text{poly}(n, 2^k)$)
 - **Result is tight for** $k \geq \log(n)^{2.1}$ (2^k is a lower bound)
- Recall: Degree of Fourier representation of g
 - [v-z-Gathen Roche]: $\deg(g) = k - k^{0.525}$
 - Implies our result when $\hat{g}(\emptyset) = 0$

Proof

Main theorem

- **Q:** what is the degree of minimal term in Fourier spectrum of g (if g is not Parity)
- **Theorem:** g symmetric nonlinear Boolean function. Then, there exists $|S| \leq k^{0.525}$ such that $\hat{g}(S) \neq 0$.
- **Proof:** First, something NOT completely different!

Two easy facts

- **Fact:** deg of Fourier spectrum of $g = \deg(g_{\mathbb{N}})$
- **Proof:** lin. Trans. does not affect degree
- **Fact:** the polynomials

$$\binom{x}{m} = \frac{x(x-1)(x-2)\dots(x-m+1)}{m!}$$

form a basis to the space of polynomials over integers, with integer coefficients:

$$g_{\mathbb{N}}(|x|) = \sum a_i \binom{|x|}{i} \text{ where } a_i \text{ are integers}$$

Degree of symmetric functions

- [v-z-Gathen Roche]: $\deg(g) \geq k - k^{0.525}$
- **Fact:** $g_{\mathbb{N}}(|x|) = \sum_{i=0}^k a_i \binom{|x|}{i}$ where a_i integers
- Proof [vzG-R]: assume $k=p-1$, p prime
- Set $f_i(j) \equiv_p 1$ iff $j=i$ i.e., $f_i(x) = 1 - (x-i)^{p-1}$
- $\Rightarrow g_{\mathbb{N}}(x) \equiv_p \sum f_i(x)g_{\mathbb{N}}(i)$
- Coefficient of x^{p-1} is $-\sum g_{\mathbb{N}}(i) \neq 0$ (g not const.)
- Taking mod p can only reduce degree
- $\Rightarrow \deg(g_{\mathbb{N}}) = p-1 = k$
- Now use density of primes

Degree of symmetric functions

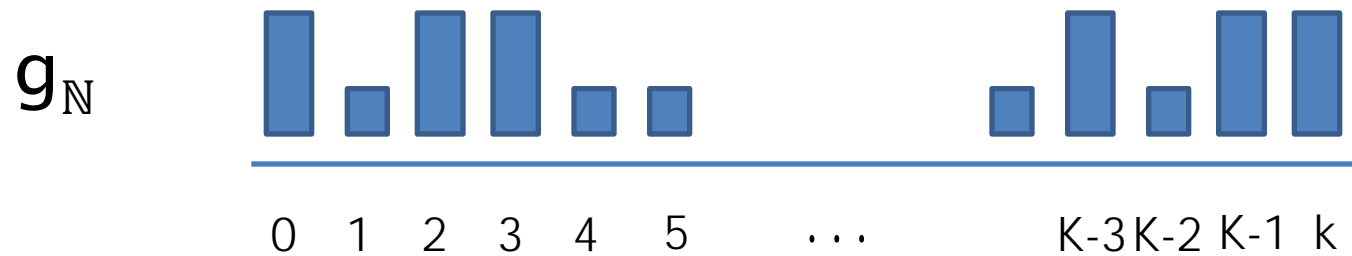
- [v-z-Gathen Roche]: $\deg(g) \geq k - k^{0.525}$
- **Fact:** $g_{\mathbb{N}}(|x|) = \sum_{i=0}^k a_i \binom{|x|}{i}$ where a_i integers
- Proof
- Set $S = [k - k^{0.525}, k]$
 Reason for mysterious 0.525:
 always exists prime number in $[k - k^{0.525}, k]$
- \Rightarrow ERH: in $[k - k^{0.5}, k]$
- Coefficient of x^k is $\sum a_i \binom{k}{i} / k!$ (g not const.)
- Taking mod p can only reduce degree
- $\Rightarrow \deg(g_{\mathbb{N}}) = p-1 = k$
- Now use density of primes

Proof overview

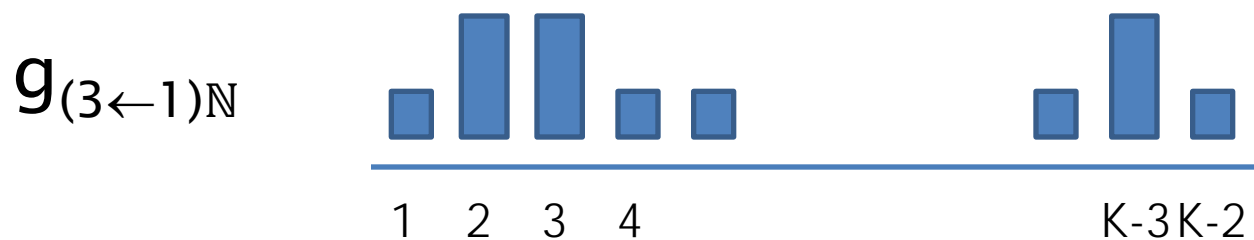
- Starting point: $\hat{g}(S) = 0$ for all $0 \neq |S| \leq t$ iff fixing t variables does not affect $\Pr[g=1]$
- Express $\Pr[g=1]$, for every restriction of t variables, as linear combination of values of g (weights being binomial coefficients)
- Deduce from equations that g is fixed on large intervals, or $g = \text{Parity}$
- Conclude (in contradiction) that g has a low degree Fourier coefficient

Restrictions, bias and Fourier transform

- **Def:** $\text{bias}(g) = \Pr_x[g(x)=1]$
- **Intuition** (for learning problem) consider samples with $x_1=0$. If bias changes then x_1 in the junta!
- We shall study such restrictions of g
- **Def:** $g_{(m \leftarrow r)}$: substituting r 1's to last m variables
- **Examples:**
 - $\text{MAJ}(x_1, x_2, x_3, x_4, x_5)_{(2 \leftarrow 0)} = x_1 \wedge x_2 \wedge x_3$
 - $\text{Parity}^k_{(m \leftarrow 7)} = \text{Parity}^{k-m} \oplus 1$
- $g_{(m \leftarrow r)\mathbb{N}}(|x|) = g_{\mathbb{N}}(|x|+r)$ (but now $x=0, \dots, k-m$)



$$|x| = 0, \dots, k$$



$$|x| = 0, \dots, k-3$$

$$g_{(m \leftarrow r)N}(|x|) = g_N(|x|+r), \quad |x| = 0, \dots, k-m$$

Restrictions, bias and Fourier transform

- Recall:

- $\text{bias}(g) = \Pr_x[g(x)=1]$
- $\hat{g}(S) =$ Fourier coefficient of g at S
- $g_{(m \leftarrow r)}$ = substituting r 1's to last m variables

- Thm [Xiao Massey]: Following are equivalent

- $\hat{g}(S) = 0$ for all $0 \neq |S| \leq t$
- For all $0 \leq r \leq m \leq t$: $\text{bias}(g_{(m \leftarrow r)}) = \text{bias}(g)$
(i.e. substitutions don't affect bias)

- Proof: Consider the Fourier transform. It is easy to see that... (both directions are the easy direction)

- From now on: focus on bias of $g_{(m \leftarrow r)}$

More on the bias of g

- Recall representation as $g_{\mathbb{N}}: \{0, \dots, k\} \rightarrow \{0, 1\}$
- $\text{bias}(g) = \Pr[g(x)=1] = \frac{1}{2^k} \sum_{i=0}^k \binom{k}{i} g_{\mathbb{N}}(i)$
- $g_{(m \leftarrow r)\mathbb{N}}(x) = g_{\mathbb{N}}(x+r)$ ($x=0, \dots, k-m$)
- $\text{bias}(g_{(m \leftarrow r)}) = \frac{1}{2^{k-m}} \sum_{i=0}^{k-m} \binom{k-m}{i} g_{\mathbb{N}}(i+r)$
- **Observation:** If $\text{bias}(g_{(m \leftarrow r)}) = \text{bias}(g)$ then we get **many** linear equations in $\{g_{\mathbb{N}}(i)\}_{i=0, \dots, k}$
- **Main obstacle:** how to use this information: coefficients are complicated, solutions are 0/1

Going modular

- Insight from [von zur Gathen Roche]: study equations modulo prime numbers

- Main tool: Lucas' theorem

Let $a = \sum a_i \cdot p^i$ and $b = \sum b_i \cdot p^i$ then $\binom{a}{b} \equiv_p \prod_i \binom{a_i}{b_i}$

- Examples:

- $0 < r < p$ then $\binom{p}{r} \equiv_p \binom{1}{0} \binom{0}{r} = 0$

- $\binom{2p}{r} \equiv_p \binom{2}{r_1} \binom{0}{r_0}$ not zero only for $r \in \{0, p, 2p\}$

- $0 < r < p$ then $\binom{p-1}{r} \equiv_p (-1)^r$

- **The point is:** it helps simplify equations

Bias goes modular

- Let $p \sim k - k^{0.525}$ prime, $t = k - (p-1) \sim k^{0.525}$
after fixing t vars, $p-1$ vars left
- For every $0 \leq r \leq t$ we have $\text{bias}(g) = \text{bias}(g_{(t \leftarrow r)})$

$$\text{bias}(g) = \text{bias}(g_{(t \leftarrow r)}) = \frac{1}{2^{p-1}} \sum_{i=0}^{p-1} \binom{p-1}{i} g_{\mathbb{N}}(i+r)$$

$$2^{p-1} \cdot \text{bias}(g) = \sum_{i=0}^{p-1} \binom{p-1}{i} g_{\mathbb{N}}(i+r) \equiv_p \sum_{i=0}^{p-1} (-1)^i g_{\mathbb{N}}(i+r)$$

Adding the equations for r and $r+1$ we get

$$2 \cdot 2^{p-1} \cdot \text{bias}(g) \equiv_p g(r) + g(p+r)$$

$$\Rightarrow 2 \cdot 2^{p-1} \cdot \text{bias}(g) \in_p \{0, 1, 2\}$$

$$\Rightarrow 2^{p-1} \cdot \text{bias}(g) \in_p \{0, 1, \frac{1}{2}(p+1)\}$$

More on the values of g

• **Summary:** if $\forall |S| \leq t \sim k^{0.525} \hat{g}(S)=0$ then

– $2 \cdot 2^{p-1} \cdot \text{bias}(g) \equiv_p g(r) + g(p+r)$

– $2^{p-1} \cdot \text{bias}(g) \in_p \{0, 1, \frac{1}{2}(p+1)\}$

• **Case $2^{p-1} \cdot \text{bias}(g) \equiv_p \frac{1}{2}(p+1)$:** then $\forall 0 \leq r \leq t$

$$\frac{1}{2}(p+1) = 2^{p-1} \cdot \text{bias}(g) \equiv_p \sum_{i=0}^{p-1} (-1)^r g_{\mathbb{N}}(i+r)$$

$$= (g_{\mathbb{N}}(r) + g_{\mathbb{N}}(2+r) + g_{\mathbb{N}}(4+r) + \dots) - (g_{\mathbb{N}}(1+r) + g_{\mathbb{N}}(3+r) + \dots)$$

$\Rightarrow g$ is **Parity!** (or \neg Parity)

• **Case $2^{p-1} \cdot \text{bias}(g) \in_p \{0, 1\}$:** $\forall 0 \leq r \leq t \quad g(r) = g(p+r) = c_p$

$\Rightarrow g$ is fixed on small and large inputs

Conclusion: $g \neq \text{Parity}$ then $\forall 0 \leq r \leq t \quad g_{\mathbb{N}}(r) = g_{\mathbb{N}}(p+r) = c_p$

Even more on the values of g

- **Summary:** if $\forall |S| \leq t \sim k^{0.525} \quad \hat{g}(S)=0$ then
 - g is Parity or
 - $\forall 0 \leq r \leq t = k - p + 1 \quad g(r) = g(p+r) = c_p$
 - $g_{\mathbb{N}}$ fixed on $[0, k^{0.525}] \cup [k - k^{0.525}, k]$
- Next: $g_{\mathbb{N}}(x)$ fixed at $x \in [\frac{1}{2}k - k^{0.525}, \frac{1}{2}k + k^{0.525}]$
- Take prime $q \sim \frac{1}{2}k - k^{0.525}$ and study $g_{(k-2q \leftarrow r)\mathbb{N}}$

$$2^{2q} \cdot \text{bias}(g) = \sum_{i=0}^{2q} \binom{2q}{i} g_{\mathbb{N}}(i+r) \equiv_q$$

$$g_{\mathbb{N}}(r) + 2g_{\mathbb{N}}(q+r) + g_{\mathbb{N}}(2q+r) = 2g_{\mathbb{N}}(q+r) + 2c_p$$
- **Lesson:** $\forall 0 \leq r \leq k - 2q \quad g_{\mathbb{N}}(q+r) = c_q$

Concluding the proof

- **Summary:** if $\forall |S| \leq t \sim k^{0.525} \quad \hat{g}(S)=0$ then
 - g is Parity or
 - $g_{\mathbb{N}}(x)$ is fixed at $x \in [\frac{1}{2}k - k^{0.525}, \frac{1}{2}k + k^{0.525}]$ (wlog $g_{\mathbb{N}}=0$ there)
- **Claim:** if $g_{\mathbb{N}}(x)$ is fixed at $x \in [\frac{1}{2}k - k^{0.525}, \frac{1}{2}k + k^{0.525}]$ then $\text{bias}(g_{\mathbb{N}}) \neq \text{bias}(g_{(2 \leftarrow 1)\mathbb{N}})$
- **Proof:** $\frac{1}{2^k} \binom{k}{i} \leq \frac{1}{2^{k-2}} \binom{k-2}{i-1}$ iff $i \in [\frac{1}{2}k - k^{0.5}, \frac{1}{2}k + k^{0.5}]$
 Contribution of $g_{\mathbb{N}}(i)$ to bias always larger in $g_{\mathbb{N}}$:

$$\text{bias}(g) = \frac{1}{2^k} \sum_{i=0}^k \binom{k}{i} g_{\mathbb{N}}(i)$$

$$\text{bias}(g_{(2 \leftarrow 1)\mathbb{N}}) = \frac{1}{2^{k-2}} \sum_{i=0}^{k-2} \binom{k-2}{i} g_{\mathbb{N}}(i+1)$$

Recap

- **Q:** what is the degree of minimal term in Fourier spectrum of g (if g is not Parity)
- $\text{bias}(g) = \frac{1}{2^k} \sum_{i=0}^k \binom{k}{i} g_{\mathbb{N}}(i)$
- **Thm:** $\forall |S| \leq t \hat{g}(S) = 0$ iff $\forall 0 \leq r \leq m \leq t \text{ bias}(g_{(m \leftarrow r)}) = \text{bias}(g)$
- Studied bias modulo different primes p, q after substituting values to $k - (p-1)$, $k - 2q$ variables
- Found a lot of information on values of g and concluded $g_{\mathbb{N}}(i)$ fixed for i 's around $\frac{1}{2}k$
- This implied that $\text{bias}(g_{(2 \leftarrow 1)}) \neq \text{bias}(g) \perp$

How to improve?

- **Theorem:** If the degree of any polynomial $h:\{0,\dots,k-2\} \rightarrow \{0,1,2\}$ is at least $k-s$, then for any symmetric (non linear) g on $\{0,1\}^k$ there exists $|S|<s$ satisfying $\hat{g}(S)\neq 0$
- Motivates the study of degree of polynomials $h:\{0,\dots,k\} \rightarrow \{0,1,2\}$
- Currently best known is [**Cohen-S-Tal**]
 $\deg(h) \geq k - k/\log\log(k)$
- **Open questions:** obvious...

Thanks for listening