

Linial-Mansour-Nisan Theorem

Crash Course on Fourier Analysis on Boolean Functions

Edward Kim

November 12, 2020

1 Fourier Analysis on Boolean Functions

- Characters
- The Fourier Transform

2 LMN Theorem

3 Consequences

- Approximation by Low-degree Polynomials
- Sensitivity and Influence

Let us first define some general concepts for any finite abelian group G .

Definition

A group homomorphism from $\chi : G \rightarrow \mathbb{C}^*$ is called a *character* of G where $\mathbb{C}^* = \mathbb{C}/\{0\}$

- 1 We call the homomorphism $\chi_0 : G \rightarrow \mathbb{C}^*$, $\chi_0 = 1$ the *trivial character* of G .
- 2 $\chi(a + b) = \chi(a)\chi(b)$
- 3 The characters of G form an abelian group \hat{G} under pointwise multiplication of complex-valued functions. The group \hat{G} is known as the *character group* of G .

Characters

- 1 It follows from basic properties of cyclic groups that the only characters of \mathbb{Z}_n are ones of the form

$$\chi_j(x) = e^{2\pi i j x / n} \quad j \in [n], x \in \mathbb{Z}_n$$

Note that each character is associated to an element of the group $j \in \mathbb{Z}_n$

Theorem (Characters for finite abelian groups)

For any finite abelian group $G \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \cdots \times \mathbb{Z}_{n_k}$: the characters for G will be:

$$\chi_a(x) = \prod_{i \in [k]} e^{2\pi i a_i x_i / n_i}$$

for $a \in \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \cdots \times \mathbb{Z}_{n_k}$

Characters

- 1 It turns out the characters live in the vector space $L_2(G)$ i.e maps from $\phi : G \rightarrow \mathbb{C}$ square-integrable in respect to the uniform probability measure: $\frac{1}{|G|} \sum_{x \in G} |\phi(x)|^2 < \infty$
- 2 Actually a Hilbert space endowed with the inner product:

$$\langle f, g \rangle = \frac{1}{|G|} \sum_{x \in G} f(x) \overline{g(x)} = \mathbb{E}_x f(x) \overline{g(x)}$$

Theorem

The characters $\chi \in \hat{G}$ form an orthonormal basis for $L_2(G)$

- 3 With any orthonormal basis, one can decompose any vector into its direct sum decomposition.

Characters

- 1 We restrict our attention to the finite abelian group \mathbb{Z}_2^n as a natural group to define boolean functions $f : \{0, 1\}^n \rightarrow \{0, 1\}$.
- 2 By the observations above, we conclude that our characters for $\mathbb{Z}_2^n = \mathbb{Z}_2 \times \dots \times \mathbb{Z}_2$ are defined as

$$\chi_a(x) = (-1)^{\sum_{i \in [n]} x_i a_i} = (-1)^{\sum_{i \in [n], a_i=1} x_i}$$

where x_i is the i^{th} bit of $x \in \{0, 1\}^n$.

Some Notation

Let $S \subseteq [n]$ be such that $S_x = \{i \mid x_i = 1\}$. As there is a bijective correspondence between all such subsets S_x and $x \in \{0, 1\}^n$, we will sometimes identify bit strings with their subset counterparts

$$\chi_A(x) = (-1)^{\sum_{i \in A} x_i}, \quad A \subseteq [n]$$

The Fourier Transform

- 1 Given any $f \in L_2(G)$, let $\hat{f} : \hat{G} \rightarrow \mathbb{C}$ be the complex-valued function such that

$$\hat{f}(\chi_a) = \langle f, \chi_a \rangle, \quad a \in G$$

These are the projections onto the orthonormal basis of characters. We deem \hat{f} the *Fourier Transform* of f .

- 2 The direct sum decomposition of f yields the following form known as the *Fourier Inversion Formula*:

$$f = \sum_{a \in G} \hat{f}(\chi_a) \chi_a$$

The complex value $\hat{f}(\chi_a)$ is called the *Fourier coefficient* associated to χ_a

- 3 From the above definition, we directly calculate that

$$\hat{f}(0) = \langle f, \chi_0 \rangle = \mathbb{E}_x f(x)$$

The Fourier Transform

1 Let $f = \text{Maj}_3$ where $G = \mathbb{Z}_2^n$.

$$\hat{f}(0^3) = \mathbb{E}_x[f(x)] = \frac{1}{2} \quad (1)$$

$$\hat{f}(\{001, 010, 100\}) = -\frac{1}{4} \quad (2)$$

$$\hat{f}(\{011, 110, 101\}) = 0 \quad (3)$$

$$\hat{f}(1^3) = \frac{1}{8} \sum_{x \in \{0,1\}^3} (-1)^{|x|} f(x) = \frac{1}{4} \quad (4)$$

Definition

Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be an n -ary boolean function. The *Fourier degree* of f , denoted by $\text{def}_{\mathcal{F}}(f)$, is the largest $|S|$ such that $\hat{f}(S) \neq 0$.

For the case of $f = \text{Maj}_3$, $\text{def}_{\mathcal{F}}(f) = 3$

The Fourier Transform

Theorem (Parseval's Identity)

Let $f \in L_2(G)$. Then

$$\|f\|_2^2 = \sum_{a \in G} |\hat{f}(a)|^2$$

Proof.

$$\begin{aligned} \|f\|_2^2 = \langle f, f \rangle &= \left\langle \sum_{a \in G} \hat{f}(a) \chi_a, \sum_{b \in G} \hat{f}(b) \chi_b \right\rangle = \sum_{a, b \in G} \hat{f}(a) \overline{\hat{f}(b)} \langle \chi_a, \chi_b \rangle \\ &= \sum_{a \in G} |\hat{f}(a)|^2 \end{aligned}$$

The last equality is just the orthonormality of the characters. □

Theorem

(Linial, Mansour, Nisan) Let f be a boolean function computed by a circuit of depth d and size M and let t be any non-negative integer.

Then

$$\sum_{|S| > t} |\hat{f}(S)|^2 \leq 2M2^{-t^{1/d}/20} \quad (5)$$

The theorem reveals that the t -tails of the Fourier spectrum, i.e strings indexed by sets $|S| > t$, become exponentially small in t for boolean functions in AC^0 .

LMN Theorem

Theorem (*Håstad*)

Let f be given by a CNF-formula where each clause has size at most t , and choose a random restriction ρ with parameter p such that $\Pr[\rho(x_i)] = p$ for all input variables x_i . With probability of at least $1 - (5pt)^s$, f_ρ can be expressed as a DNF formula where each clause has size of at most s , and the clause all accept disjoint sets of inputs i.e no string $x \in \{0,1\}^n$ satisfies more than one clause.

Corollary

Let f be a boolean function computed by a CNF of bottom fan-in of at most t , and ρ is a p -random restriction, then

$$\Pr[\deg_{\mathcal{F}}(f_\rho) > s] < (5pt)^s \quad (6)$$

Corollary (Tail Degree Corollary)

Let f be a boolean function computed by a circuit of size M and depth d . Then

$$\Pr[\deg(f_\rho) > s] \leq M2^{-s}$$

where ρ is a random restriction where $p = \frac{1}{10^d s^{d-1}}$

LMN Theorem

Corollary (Tail Degree Corollary)

Let f be a boolean function computed by a circuit of size M and depth d . Then

$$\Pr[\deg(f_\rho) > s] \leq M2^{-s}$$

where ρ is a random restriction where $p = \frac{1}{10^d s^{d-1}}$

Proof Sketch.

Show that first random restriction of parameter $p_0 = \frac{1}{10}$, the bottom gates' fan-ins are at most s with probability of at least $1 - 2^{-s}$. Then iterate Håstad's switching lemma with under $p_i = \frac{1}{10^s}$ on each gate of distance two from the input variables to turn them into DNFs with disjoint inputs. Collapse a level and the lemma ensures that the new bottom fan-in is at most s . Stop when we are left with a CNF (depth-2) with bottom fan-in of at most s and invoke the previous corollary. \square

Proof of LMN Theorem

Definition

Let k be a positive integer and $f : \mathbb{Z}_2^n \rightarrow \mathbb{C}$ a complex-valued function on \mathbb{Z}_2^n . We define:

$$f^{\leq k} := \sum_{|S| \leq k} \hat{f}(S) \chi_S$$

The notation symbols $f^{=k}, f^{\geq k}$ are defined in the same manner.

Proof of LMN Theorem

Definition

Let k be a positive integer and $f : \mathbb{Z}_2^n \rightarrow \mathbb{C}$ a complex-valued function on \mathbb{Z}_2^n . We define:

$$f^{\leq k} := \sum_{|S| \leq k} \hat{f}(S) \chi_S$$

The notation symbols $f^{=k}, f^{\geq k}$ are defined in the same manner.

- 1 We begin by first setting our probability parameter $p \leq \frac{1}{10^d k^{d-1}}$. The values p, k will be fixed later to invoke the Tail Degree Corollary above.

Proof of LMN Theorem

Definition

Let k be a positive integer and $f : \mathbb{Z}_2^n \rightarrow \mathbb{C}$ a complex-valued function on \mathbb{Z}_2^n . We define:

$$f^{\leq k} := \sum_{|S| \leq k} \hat{f}(S) \chi_S$$

The notation symbols $f^{\leq k}, f^{\geq k}$ are defined in the same manner.

- 1 We begin by first setting our probability parameter $p \leq \frac{1}{10^d k^{d-1}}$. The values p, k will be fixed later to invoke the Tail Degree Corollary above.
- 2 Recall how we sample a random restriction ρ by sampling some $V \subseteq [n]$ to be the indices which are *not* set to $*$. Each index has a $1 - p$ probability of being set to either 0, 1. For each index in V , we uniformly sample some bit string in $\{0, 1\}^{|V|}$ to fix the indices contained in V .

Proof of LMN Theorem

- 1 Let x_V be the restriction of string x on the input indices to those found in V

Proof of LMN Theorem

- 1 Let x_V be the restriction of string x on the input indices to those found in V
- 2 If we recall that our characters are defined as $\chi_A(x) = (-1)^{\sum_{i \in A} x_i}$

Proof of LMN Theorem

- 1 Let x_V be the restriction of string x on the input indices to those found in V
- 2 If we recall that our characters are defined as $\chi_A(x) = (-1)^{\sum_{i \in A} x_i}$

How characters separate

$$\begin{aligned}\chi_S(x) &= (-1)^{\sum_{i \in S} x_i} = (-1)^{\sum_{i \in S \cap V} x_i + \sum_{i \in S/V} x_i} \\ &= (-1)^{\sum_{i \in S \cap V} x_i} (-1)^{\sum_{i \in S/V} x_i} = \chi_{S \cap V}(x_V) \chi_{S/V}(x_{\overline{V}})\end{aligned}$$

Proof of LMN Theorem

- 1 If we set x_V to some bit string, it makes sense to think of $f_{x_V} = f(x_V, *)$ as a function $f_{x_V} : \{0, 1\}^{|x_{\overline{V}}|} \rightarrow \{0, 1\}$.

Proof of LMN Theorem

- 1 If we set x_V to some bit string, it makes sense to think of $f_{x_V} = f(x_V, *)$ as a function $f_{x_V} : \{0, 1\}^{|x_{\overline{V}}|} \rightarrow \{0, 1\}$.
- 2 By our Fourier Inversion Formula and the observation in the previous slide:

$$\begin{aligned} f(x) &= \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x) = \sum_{V \sqcup \overline{V} \subseteq [n]} \hat{f}(S) \chi_{S \cap V}(x_V) \chi_{S/V}(x_{\overline{V}}) \\ &= \sum_{H \subseteq \overline{V}} \left(\sum_{J \subseteq V} \hat{f}(H \cup J) \chi_J(x_V) \right) \chi_H(x_{\overline{V}}) \end{aligned}$$

- 3 Since the Fourier decomposition is unique for f_{x_V} ,

$$\widehat{f_{x_V}}(H) = \sum_{J \subseteq V} \hat{f}(H \cup J) \chi_J(x_V), \quad H \subseteq \overline{V}$$

Proof of LMN Theorem

1 By Parseval's identity on the function $x_T \mapsto \widehat{f_{x_T}}(H)$,

$$\mathbb{E}_{x_V} |\widehat{f_{x_V}}(H)|^2 = \langle \hat{f}_-(H), \hat{f}_-(H) \rangle = \sum_{J \subseteq V} |\hat{f}(H \cup J)|^2 \quad (7)$$

Proof of LMN Theorem

- 1 By Parseval's identity on the function $x_T \mapsto \widehat{f_{x_T}}(H)$,

$$\mathbb{E}_{x_V} |\widehat{f_{x_V}}(H)|^2 = \langle \hat{f}_-(H), \hat{f}_-(H) \rangle = \sum_{J \subseteq V} |\hat{f}(H \cup J)|^2 \quad (7)$$

- 2 This along with another application of Parseval's identity yields the string of equalities:

$$\mathbb{E}_{x_V} \|f_{x_V}^{>k}\|_2^2 = \mathbb{E}_{x_V} \sum_{\substack{H \subseteq \bar{V} \\ |H| > k}} |\widehat{f_{x_V}}(H)|^2 = \sum_{\substack{H \subseteq \bar{V} \\ |H| > k}} \sum_{J \subseteq V} |\hat{f}(H \cup J)|^2 \quad (8)$$

$$= \sum_{\substack{S \subseteq [n] \\ |S \cap \bar{V}| > k}} |\hat{f}(S)|^2 \quad (9)$$

Proof of LMN Theorem

- 1 Sampling over all such p -restrictions further shows that the lefthand side is upper bounded by:

$$\mathbb{E}_V \mathbb{E}_{x_V} \|f_{x_V}^{>k}\|_2^2 = \mathbb{E}_\rho \|f_\rho^{>k}\|_2^2 \leq \Pr[\deg_{\mathcal{F}}(f_\rho) > k] \leq M2^{-k}$$

Proof of LMN Theorem

- 1 Sampling over all such p -restrictions further shows that the lefthand side is upper bounded by:

$$\mathbb{E}_V \mathbb{E}_{x_V} \|f_{x_V}^{>k}\|_2^2 = \mathbb{E}_\rho \|f_\rho^{>k}\|_2^2 \leq \Pr[\deg_{\mathcal{F}}(f_\rho) > k] \leq M2^{-k}$$

- 2 A Chernoff bound argument shows that the righthand side is lower bounded by:

$$\sum_{|S|>t} |\hat{f}(S)|^2 \leq 2\mathbb{E}_T \sum_{\substack{S \subseteq [n] \\ |S \cap \bar{V}| > k}} |\hat{f}(S)|^2$$

for $k = pt/2$

Proof of LMN Theorem

- 1 Sampling over all such p -restrictions further shows that the lefthand side is upper bounded by:

$$\mathbb{E}_V \mathbb{E}_{x_V} \|f_{x_V}^{>k}\|_2^2 = \mathbb{E}_\rho \|f_\rho^{>k}\|_2^2 \leq \Pr[\deg_{\mathcal{F}}(f_\rho) > k] \leq M2^{-k}$$

- 2 A Chernoff bound argument shows that the righthand side is lower bounded by:

$$\sum_{|S|>t} |\hat{f}(S)|^2 \leq 2\mathbb{E}_T \sum_{\substack{S \subseteq [n] \\ |S \cap \bar{V}| > k}} |\hat{f}(S)|^2$$

for $k = pt/2$

- 3 Setting our constants $p = \frac{1}{10t^{(d-1)/d}}$ and $k = t^{1/d}/20$ shows that $p \leq \frac{1}{10d k^{d-1}}$. This allows us to invoke our Tail Degree Corollary which gives us the desired inequality

$$\sum_{|S|>t} |\hat{f}(S)|^2 \leq 2M2^{-t^{1/d}/20} \quad (10)$$

Approximation by Low-degree Polynomials

Theorem (6) shows that we can approximate functions in $f \in \text{AC}^0$ by taking our polynomials to be $f^{\leq k}$ for some sufficiently large k . Here we make the distinction that our Fourier expansion of f will be

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

as a polynomial with complex coefficients.

Lemma

Let $f \in \text{AC}^0$ be a boolean function of polynomial size and depth d . Then there exists a complex polynomial of degree $\mathcal{O}((\log n/\epsilon)^d)$ such that $\|f - p\|_2 < \epsilon$

Sensitivity and Influence

LMN also shows that functions in AC^0 have low average sensitivity i.e its output is not very sensitive to changes to the input.

Definition

Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a boolean function. We define the *sensitivity* of an input $x \in \{0, 1\}^n$ in respect to f , $s_f(x)$ to be the number of indices i such that $f(x) \neq f(x + e_i)$ where e_i is the bit string with zeros everywhere except for the i^{th} index.

Definition

Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a boolean function. The *influence* of f , I_f is defined as the average sensitivity over all input bit strings

$$I_f = \mathbb{E}_x[s_f(x)] \quad (11)$$

Sensitivity and Influence

The influence of f can be expressed in terms of its Fourier coefficients:

$$I_f = 4 \sum_{S \subseteq [n]} |S| |\hat{f}(s)|^2 \quad (12)$$

By combining this equivalence with Theorem (6), we deduce the following upper bound on the influence of a function in AC^0

Lemma

Let $f \in \text{AC}^0$ be of depth d . Then

$$I_f = \mathcal{O}((\log n)^d) \quad (13)$$

The lemma shows that functions in AC^0 are not suitable for constructing universal hash functions and pseudorandom function generators.

Sensitivity and Influence

Definition

A function $f : \{0, 1\}^m \times \{0, 1\}^n \rightarrow \{0, 1\}$ is called a *pseudorandom function generator* if there exists no polynomial-time oracle Turing machine which can distinguish between the outputs from a true random oracle versus $f(s, \cdot)$ for some random seed $s \in \{0, 1\}^m$.

Lemma

No pseudorandom function generators exist in AC^0

By taking advantage of the low average sensitivity, we can simply perturb the input slightly and check if the output of f changes. If it doesn't, there is a good chance that it lies in AC^0 .