1 Last time and today

Previously:

- Fourier concentration for DNFs and constant-depth circuits (Random restriction)
- Started Fourier concentration for monotone functions (Influence)

Today:

- Finish Fourier concentration via total influence bound for monotone functions
- Digressions (Open questions)
  - Gotsman-Linial Conjecture
  - Fourier Degree vs. Max Sensitivity

Relevant Readings:

- Gotsman-Linial Conjecture

- Fourier degree versus sensitivity
    (they refer to it as the ”block sensitivity versus sensitivity” question; as they show, the block sensitivity and the Fourier degree are polynomially related)
2 Influences

Recall the definition of influences of Boolean functions.

**Definition 1.** For any function \( f : \{-1, 1\}^n \to \{-1, 1\} \), the influence of coordinate \( i \) is defined to be the probability that \( f(x^i) \neq f(x^{i-1}) \) for a random input \( x \):

\[
\text{Inf}_i(f) = \Pr_{x \sim \{-1, 1\}^n}[f(x^i) \neq f(x^{i-1})]
\]

In addition, the total influence is defined to be

\[
\text{Inf}(f) = \sum_{i=1}^{n} \text{Inf}_i(f)
\]

**Example 2.** \( \text{Inf}(\text{MAJ}) = \Theta(\sqrt{n}) \). Imagine the majority function. It is easy to see that flopping of the function happens when the number of \( +1 \) and \( -1 \) in other coordinates is same. In other words, for \( \text{MAJ}_n \), the \( i \)th voter has influence if and only if the other \( n-1 \) votes split evenly. This event happens with probability \((\frac{n-1}{n-2})^2\), and it is asymptotically \( \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{n}} \) by Stirling’s formula. Thus,

\[
\text{Inf}(f) = \sum_{i=1}^{n} \text{Inf}_i(f) \approx \sqrt{\frac{2}{\pi}} \sqrt{n}.
\]

**Definition 3.** For any function \( f : \{-1, 1\}^n \to \mathbb{R} \), \( i \)th difference operator maps the function \( f \) to

\[
D_i f(x) = \frac{f(x^i) - f(x^{i-1})}{2}.
\]

Note that for Boolean function \( f \)

\[
D_i f(x) = \begin{cases} 
0 & \text{if } f \text{ is insensitive to variable } i \text{ on input } x, \text{ i.e. } f(x^i) = f(x^{i-1}) \\
\pm 1 & \text{otherwise.}
\end{cases}
\]

Therefore,

\[
E[(D_i f(x))^2] = \Pr[f(x^i) \neq f(x^{i-1})] = \text{Inf}_i(f).
\]

\(^1\)Note that there is another definition \( \text{Inf}_i(f) = \Pr_{x \sim \{-1, 1\}^n}[f(x) \neq f(x^i)] \). They are exactly same.
Proposition 4. For any function \( f : \{-1, 1\}^n \to \{-1, 1\}^n \),
\[
\text{Inf}_i(f) = \sum_{S \ni i} \hat{f}(S)^2.
\]

Proof.
\[
D_i f(x) = \left( \sum_S \hat{f}(S) \chi_S(x^{i-1}) - \sum_S \hat{f}(S) \chi_S(x^{i+1}) \right)/2 \quad \text{by definition}
\]
\[
= \sum_{S \ni i} \hat{f}(S) \chi_{S \setminus \{i\}}(x).
\]
See an example for the last equality. when \( i = 1, S = \{2, 3\} \):
\[
\hat{f}(2, 3)x_2x_3 - \hat{f}(2, 3)x_2x_3 = 0
\]
when \( i = 1, S = \{1, 2\} \):
\[
\hat{f}(1, 2)x_1 - \hat{f}(1, 2)(-1)x_2 = 2\hat{f}(1, 2) \cdot x_2.
\]
By generalizing above, we have only Fourier coefficients which include \( i \). So
\[
\text{Inf}_i(f) = E[(D_i(f))^2]
\]
\[
= \sum_{S \ni i} \hat{f}(S)^2. \quad \text{(by Parseval)}
\]

From the above proposition, it is straightforward to see
\[
\text{Inf}(f) = \sum_S |S| \hat{f}(S)^2.
\]
Also, \( \hat{f}(S)^2 \) are nonnegative and \( \sum_S \hat{f}(S)^2 = 1 \) (Parseval’s identity) for Boolean functions. So we can view \( \hat{f}(S)^2 \) as defining \( D \) distribution over \( S \), and view \( \text{Inf}(f) \) as \( E_{S \sim D}[|S|] \), “average degree”.

Observation 5. If \( f \) is balanced, i.e. \( E[f] = 0 \), then \( \text{Inf}(f) \geq 1 \). (Note that \( E[f] = \hat{f}(0) \)
2 INFLUENCES

Proof.

\[ \text{Inf}(f) = \sum_S |S| \hat{f}(S)^2 \]
\[ = \hat{f}(\emptyset)^2 \cdot 0 + \sum_{|S| \geq 1} |S| \hat{f}(S)^2 \]
\[ \geq \sum_{|S| \geq 1} \hat{f}(S)^2 = 1 \] (by Parseval’s identity)

\[ \square \]

2.1 Monotone Functions

Theorem 6. Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \) and monotone. Then \( \text{Inf}_i(f) = \hat{f}(i) \).

(\( \hat{f}(i) = \hat{f}(\{i\}) \))

Proof. Consider the Fourier coefficient of a coordinate \( i \):

\[ \hat{f}(i) = E[f(x)x_i] = \Pr[f(x) = x_i] - \Pr[f(x) \neq x_i]. \]

Without loss of generality, suppose \( i = 1 \). Then,

\[ \text{Inf}_1(f) = \Pr_x[f(x^i \leftarrow 1) \neq f(x^i \leftarrow -1)] \]
\[ = \Pr_{x_2, \ldots, x_n}[f(1, x_2, \ldots, x_n) \neq f(-1, x_2, \ldots, x_n)] \]
\[ = \frac{1}{2^{n-1}} \cdot \text{[the number of } x' = (x_2, \ldots, x_n) \text{ s.t. } f(1, x') = +1 \& f(-1, x') = -1]\] (by monotonicity)

Now go back to the Fourier coefficient of the coordinate 1.

\[ \hat{f}(1) = E[f(x)x_1] \]
\[ = \frac{1}{2^n} \sum_{x \in \{-1, 1\}^n} f(x)x_1 \]
\[ = \frac{1}{2^n} \sum_{x' \in \{-1, 1\}^{n-1}} [f(1, x') - f(-1, x')] \]
\[ = \frac{1}{2^n} \cdot 2 \cdot \text{[the number of } x' = (x_2, \ldots, x_n) \text{ s.t. } f(1, x') = +1 \& f(-1, x') = -1]\] (by monotonicity)

Therefore,

\[ \text{Inf}_i(f) = \hat{f}(i). \]

\[ \square \]
Now we use this to show that majority function has the maximum influence among all monotone Boolean functions.

**Claim 7.** For $f : \{−1, 1\}^n \to \{−1, 1\}$ and $f$ is monotone, $\text{Inf}(f) \leq \text{Inf}(\text{MAJ}) \leq \sqrt{n}$.

**Proof.** We will show that every Boolean function $f$ satisfies

$$\sum_{i=1}^{n} \hat{f}(i) \leq \sum_{i=1}^{n} \text{MAJ}(i).$$

Consider

$$\sum_{i=1}^{n} \hat{f}(i) = \sum_{i=1}^{n} E[f(x)x_i]
= E_x[f(x)(x_1 + \ldots + x_n)]
= \frac{1}{2^n} \sum_{x} f(x) \cdot (x_1 + \ldots + x_n)
\leq \frac{1}{2^n} \sum_{x} \text{sign}(x_1 + \ldots + x_n) \cdot (x_1 + \ldots + x_n),$$

and $\text{sign}(x_1 + \ldots + x_n) = \text{MAJ}(x)$. $\square$

There is an alternate proof which uses Cauchy-Schwarz inequality.

**Proof.** Fix $f$ to be some monotone function. Then,

$$\text{Inf}(f) = \sum_{i=1}^{n} \hat{f}(i).$$

We know that $\sum_{i=1}^{n} \hat{f}(i)^2 \leq \sum_{S \subseteq [n]} \hat{f}(S)^2 = 1$. Thus, $\sqrt{\sum_{i=1}^{n} \hat{f}(i)^2} \leq 1$. Cauchy-Schwarz inequality says that for any vector $u, v$, $u \cdot v \leq \|u\| \cdot \|v\|$. Let $u = (\hat{f}(1), \ldots, \hat{f}(n))$ and $v = (1, \ldots, 1)$, apply Cauchy-Schwarz inequality. Then, we get

$$\sum_{i=1}^{n} \hat{f}(i) \leq \sqrt{n}. \square$$
2.2 Influence and Fourier Concentration

Now we state a theorem which relates influence and Fourier concentration.

**Theorem 8.** Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ be any function. Then,

$$\sum_{|S| > \frac{\text{Inf}(f)}{\epsilon}} \hat{f}(S)^2 \leq \epsilon.$$  

This theorem implies that we can learn all "$\text{Inf}(f) \leq k$" functions in $n^{k/\epsilon}$ time.

**Proof.** Suppose that $\sum_{|S| > \frac{\text{Inf}(f)}{\epsilon}} \hat{f}(S)^2 > \epsilon$. Then we have

$$\text{Inf}(f) = \sum_S |S| \hat{f}(S)^2 \geq \sum_{|S| > \frac{\text{Inf}(f)}{\epsilon}} |S| \cdot \hat{f}(S)^2 > \frac{\text{Inf}(f)}{\epsilon} \cdot \sum_{|S| > \frac{\text{Inf}(f)}{\epsilon}} \hat{f}(S)^2 > \text{Inf}(f)$$

(by assumption)

which is a contradiction. \qed

**Corollary 9.** For Boolean monotone functions $f$,

$$\sum_{|S| > \frac{\sqrt{n}}{\epsilon}} \hat{f}(S)^2 \leq \epsilon.$$  

So we call learn all monotone functions in $n^{\sqrt{\frac{n}{\epsilon}}}$ time.

3 Digression: Open Question 1

Consider the influence of linear threshold functions (LTFs). Let $f$ be an LTF. Then $\text{Inf}(f)$ can be as big as $\theta(\sqrt{n})$. (MAJ) Say $f(x) = \text{sign}(a_1x_1 + \ldots + a_nx_n - \theta)$. If $a_i > 0$, then $f$ is monotonically increasing in $x_i$. If $a_i < 0$, then $f$ is monotonically decreasing in $x_i$.

**Definition 10.** A Boolean function $f$ is unate when in each coordinate $i$, $f$ is either monotonically increasing or monotonically decreasing.
It is easy to see that every LTF is unate. Here is a claim about unateness and monotonicity.

**Claim 11.** If a Boolean function $f$ is unate, then there exists a Boolean monotone function $f'$ such that $\text{Inf}_i(f) = \text{Inf}_i(f')$.

This claim is true, however we will not give a detailed proof here. Consider an example.

**Example 12.** $f(x) = \text{sign}(-3x_1 + 4x_2 - 6x_3 + 5x_4 - 11)$. This $f$ is not monotone, but unate. Then $f'(x) = \text{sign}(3x_1 + 4x_2 + 6x_3 + 5x_4 - 11)$ is monotone and $\text{Inf}_i(f) = \text{Inf}_i(f')$.

So if $f$ is a LTF, $\text{Inf}(f) \leq \sqrt{n}$. Also the Fourier concentration of LTFs is:

$$\sum_{|S| > \frac{\sqrt{n}}{4}} \hat{f}(S)^2 \leq \epsilon.$$

What about degree-$d$ polynomial threshold functions (PTFs)? Definitely not every degree-$d$ PTF is unate.

**Example 13.** Consider a degree-$2$ PTF, $\text{sign}((x_1 + \ldots + x_n)^2 - n)$. We can classify the input $x \in \{-1,1\}^n$. When $+\sqrt{n} \leq \sum x_i \leq n$, $f(x) = +1$, which includes the case $x = (1,1,\ldots,1)$. When $-\sqrt{n} < \sum x_i \leq +\sqrt{n}$, $f(x) = -1$, which includes the case $\sum_i x_i = 0$. When $-n < \sum x_i \leq -\sqrt{n}$, then $f(x) = +1$, which includes the case $x = (-1,-1,\ldots,-1)$. This implies that $f$ is not unate.

**Question 14.** What is the largest possible value of $\text{Inf}(f)$ for $f$ a degree-$d$ PTF?

Here is a conjecture for the question.

**Conjecture 15.** By [Gotsman/Linial], at most $d \cdot \sqrt{n}$.

It is known that for $d = 2$, the PTF has the influence at most $n^{3/4}$. For general $d$, the influence is at most $2^{O(d)} \cdot n^{1 - \frac{d}{4}}$.

**Conjecture 16.** Optimal degree-$d$ PTF with largest influence is the symmetric degree-$d$ PTF “slicing middle $d$ layers of hypercube”. (A symmetric degree-$d$ PTF can be expressed as $p(x_1,\ldots,x_n) = q(y)$, where $q(y)$ is a polynomial of $y = x_1 + \ldots + x_n$).
4 Application: DNFs

We can get (weaker) Fourier concentration for width-\(w\) DNFs via influence.

**Theorem 17.** Let \(f\) be a width-\(w\) DNF. Then \(\text{Inf}(f) \leq 2w\).

We need a new notion for the proof, sensitivity.

**Definition 18.** The sensitivity of \(f\) at input \(x\) is \(s_f(x) = \# \text{ of } i \in \{1, \ldots, n\} \text{ s.t. } f(x) \neq f(x^\oplus i)\). (For \(x \in \{-1,1\}^n\), \(x^\oplus i = (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n)\)) Also, the average sensitivity of \(f\) is defined to be \(\text{as}(f) = E_x[s_f(x)]\).

**Fact 19.** For any Boolean function \(f\), \(\text{Inf}(f) = \text{as}(f)\).

Now we prove the theorem.

**Proof.** Brief idea of the proof is to upper bound the number of sensitive edges \((x, y)\) in \(\{-1,1\}^n\). Let \(f = T_1 \lor \ldots \lor T_j \lor \ldots\). Suppose \(x, y \in \{-1,1\}^n\), \(y = x^\oplus i\) and \(f(x) = \text{true}, f(y) = \text{false}\). Then, \(x\) satisfies some term \(T_j\) which has less than \(w\) literals. Thus, \(s_f(x) \leq w\) because flipping any variable not in \(T_j\) will not change \(f\), i.e. \(f\) still will be “true”. This is true for all positive input \(x\). Partition all sensitive edges into buckets according to the positive input. Then, there are \(2^n\) inputs and less than \(w\) edges for each bucket, since \(s_f(x) \leq w\) for all positive input \(x\). So total number of sensitive edges in \(\{-1,1\}^n\) is less than \(2^n \cdot w\). Hence,

\[
\text{as}(f) = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} [\# \text{ of sensitive edges from } x] = \frac{1}{2^n} \cdot 2 \cdot [\text{total # of sensitive edges}] \\
\leq 2w
\]

And we know that \(\text{Inf}(f) = \text{as}(f)\). \(\square\)

**Corollary 20.** For a width-\(w\) DNF \(f\),

\[\sum_{|S| > \frac{2w}{\epsilon}} \hat{f}(S)^2 \leq \epsilon.\]
5 Digression: Open Question 2

5.1 Max Sensitivity vs. Fourier Degree

Definition 21. The sensitivity of a Boolean function $f$ is

$$s(f) = \max_{x \in \{-1, 1\}^n} s_f(x).$$

Definition 22. (Fourier) degree of a Boolean function $f$ is the degree of $f$ when viewed as a multilinear polynomial, $f(x) = \sum_S \hat{f}(S)\chi_S(s)$,

$$\deg(f) = \max |S| \text{ for all } S \text{ s.t. } \hat{f}(S) \neq 0.$$ 

Question 23. Is it always the case that $\deg(f)$ and $s(f)$ are polynomially related?

It is known that $s(f) \leq 2 \cdot \deg(f)^2$. Also, some Boolean function can have $s(f) \geq \deg(f)^{\log_2 3} \approx \deg(f)^{1.61}$.

Example 24. Consider a NAE(not-all-equal) function such that

$$\text{NAE}(x_1, x_2, x_3) = \begin{cases} +1 & \text{if } x_1 = x_2 = x_3 \\ -1 & \text{otherwise.} \end{cases}$$

We can draw a Boolean cube of NAE$(x_1, x_2, x_3)$ and see only $(+1, +1, +1)$ and $(-1, -1, -1)$ are positive inputs. Definitely, these positive inputs has the maximum sensitivity, therefore $s(\text{NAE}) = 3$. Also, the Fourier expansion of NAE is

$$\text{NAE}(x_1, x_2, x_3) = -\frac{1}{2} + \frac{1}{2}x_1x_2 + \frac{1}{2}x_1x_3 + \frac{1}{2}x_2x_3.$$ 

Thus, $\deg(\text{NAE}) = 2$. Let $h(x_1, x_2, x_3) = \text{NAE}$. Consider $f = h(h(x), h(y), h(z))$. This function has 9 variables, $\deg(f) = 4$, $s(f) = 9$. By this k-fold composition, we get $3^k$ variables, degree $2^k$ and sensitivity $3^k$.

Question 25. Is $\deg(f) \leq \text{poly}(s(f))$ for all Boolean function $f$?

We know $\deg(f) \leq 2^{O(s(f))}$ for all $f$. Also there is some simple $f$ such that $\deg(f) = n$, $s(f) = \sqrt{n}$. Maybe $\deg(f) \leq O(s(f)^2)$? (Block Sensitivity vs. Sensitivity)