CSCI-B609: A Theorist's Toolkit, Fall 2016 Nov 29

Lecture 23: Analysis of Boolean Functions

Lecturer: Yuan Zhou *Scribe: Haoyu Zhang*

1 Introduction

In this lecture, we will talk about boolean function $f : \{0,1\}^n \to \{0,1\}$ that maps n bits into a single bit. We will find that the boolean function is an abstract of all functions we use in computer science, since all computers are implemented by circuits which have on/off(0/1) two states. For our convenience, we use another represent of f, which is $f : \{-1,1\}^n \to \{-1,1\}$, and in some more general cases $f : \{-1,1\}^n \to \mathbb{R}$.

Here are 2 examples of boolean functions.

Majority function Majority function over 3 bits is defined as:

$$maj_3(x_1, x_2, x_3) = sgn(x_1 + x_2 + x_3) = \begin{cases} -1 & \text{if } x_1 + x_2 + x_3 < 0 \\ 1 & \text{otherwise} \end{cases}.$$

Parity function Parity function over 3 bits is defined as:

$$parity(x_1, x_2, x_3) = x_1 x_2 x_3.$$

Note that, we could rewrite the majority function over 3 bits as

$$\begin{aligned} maj_3(x_1, x_2, x_3) = \quad & 1 \times \left(\tfrac{1}{2} + \tfrac{x_1}{2}\right)\left(\tfrac{1}{2} + \tfrac{x_2}{2}\right)\left(\tfrac{1}{2} + \tfrac{x_3}{2}\right) \\ & +1 \times \left(\tfrac{1}{2} + \tfrac{x_1}{2}\right)\left(\tfrac{1}{2} + \tfrac{x_2}{2}\right)\left(\tfrac{1}{2} - \tfrac{x_3}{2}\right) \\ & +1 \times \left(\tfrac{1}{2} + \tfrac{x_1}{2}\right)\left(\tfrac{1}{2} - \tfrac{x_2}{2}\right)\left(\tfrac{1}{2} + \tfrac{x_3}{2}\right) \\ & -1 \times \left(\tfrac{1}{2} + \tfrac{x_1}{2}\right)\left(\tfrac{1}{2} - \tfrac{x_2}{2}\right)\left(\tfrac{1}{2} - \tfrac{x_3}{2}\right) \\ & +1 \times \left(\tfrac{1}{2} - \tfrac{x_1}{2}\right)\left(\tfrac{1}{2} + \tfrac{x_2}{2}\right)\left(\tfrac{1}{2} + \tfrac{x_3}{2}\right) \\ & -1 \times \left(\tfrac{1}{2} - \tfrac{x_1}{2}\right)\left(\tfrac{1}{2} + \tfrac{x_2}{2}\right)\left(\tfrac{1}{2} - \tfrac{x_3}{2}\right) \\ & -1 \times \left(\tfrac{1}{2} - \tfrac{x_1}{2}\right)\left(\tfrac{1}{2} - \tfrac{x_2}{2}\right)\left(\tfrac{1}{2} + \tfrac{x_3}{2}\right) \\ & -1 \times \left(\tfrac{1}{2} - \tfrac{x_1}{2}\right)\left(\tfrac{1}{2} - \tfrac{x_2}{2}\right)\left(\tfrac{1}{2} - \tfrac{x_3}{2}\right). \end{aligned}$$

Expanding it out, we have

$$maj_3(x_1, x_2, x_3) = \frac{x_1}{2} + \frac{x_2}{2} + \frac{x_3}{2} - \frac{x_1 x_2 x_3}{2}.$$

This is also called interpolation. In fact, we could express all $f : \{-1, 1\}^n \to \{-1, 1\}$ in this way, which we will discuss later.

2 Fourier Analysis for boolean functions

We first define several notations we will use later.

Definition 1. *For each subset $S \subseteq [n]$, let*

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

Definition 2. *Inner product*

$\forall\ f, g \in \{-1, 1\}^n \to \mathbb{R}$, *let*

$$\langle f, g \rangle = \mathop{\mathrm{E}}_{x \sim \{-1,1\}^n} f(x)g(x).$$

Then, we will have the following lemma.

Lemma 1. *For $S, T \subseteq [n]$, we have*

$$\langle \chi_S(x), \chi_T(x) \rangle = \begin{cases} 0 & \text{if } S \neq T \\ 1 & \text{otherwise} \end{cases}.$$

Proof. By our definition of inner product,

$$\begin{aligned} \langle \chi_S(x), \chi_T(x) \rangle &= \mathop{\mathrm{E}}_{x \sim \{-1,1\}^n} \chi_S(x)\chi_T(x) \\ &= \mathop{\mathrm{E}}_{x} \left[\prod_{i \in S} x_i \prod_{i \in T} x_i \right] \\ &= \mathop{\mathrm{E}}_{x} \left[\prod_{i \in S \cap T} x_i^2 \prod_{i \in S \triangle T} x_i \right] \\ &= \mathop{\mathrm{E}}_{x} \left[\prod_{i \in S \triangle T} x_i \right], \end{aligned}$$

where $S \triangle T$ denotes their symmetric difference.

When $S = T$, $S \triangle T = \emptyset$, so $\langle \chi_S(x), \chi_T(x) \rangle = 1$.

When $S \neq T$, $\exists j \in S \triangle T$, such that

$$\mathop{\mathrm{E}}_{x}\left[\prod_{i \in S \triangle T} x_i\right] = \mathop{\mathrm{E}}_{x}\left[\prod_{i \in S \triangle T, i \neq j} x_i\right] \mathop{\mathrm{E}}_{x_j}[x_j] = 0.$$

□

By the lemma, we know that $\chi_S(x)$ for all $S \subseteq [n]$ are orthonormal basis, and we could represent any boolean function with the basis.

Fact 1. *All functions $f \in \{-1, 1\}^n \to \mathbb{R}$ can be expressed in the following form uniquely*

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

where

$$\widehat{f}(S) = \langle f, \chi_S(x) \rangle.$$

The expression of f is called Fourier transform of f, and $\widehat{f}(S)$ is called inner Fourier transform. With the Fourier transform, we could proof the following lemma.

Lemma 2. *(Parseval Identity) For all $S, T \subseteq [n]$, we have*

$$\sum_{S \subseteq [n]} \widehat{f}(S)^2 = \mathop{\mathrm{E}}_{x}[f(x)^2] = 1.$$

Proof.

$$\begin{aligned}
\mathop{\mathrm{E}}_{x}[f(x)^2] &= \mathop{\mathrm{E}}_{x}\left[\sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x) \sum_{T \subseteq [n]} \widehat{f}(T) \chi_T(x)\right] \\
&= \mathop{\mathrm{E}}_{x}\left[\sum_{S, T \subseteq [n]} \widehat{f}(S) \widehat{f}(T) \chi_S(x) \chi_T(x)\right] \\
&= \sum_{S, T \subseteq [n]} \widehat{f}(S) \widehat{f}(T) \mathop{\mathrm{E}}_{x}[\chi_S(x) \chi_T(x)] \\
&= \sum_{S \subseteq [n]} \widehat{f}(S)^2.
\end{aligned}$$

□

Next, we talk about a new concept.

Definition 3. *The* influence of the variable x_i *on the function* $f : \{-1,1\}^n \to \{-1,1\}$ *is*

$$\mathrm{Inf}_i(f) = Pr_x[f(x) \neq f(x^i)],$$

where the x^i *denotes the vector after a value change on the* i*-th bit of* x*.*

Here are some examples: $\mathrm{Inf}_i(maj) = \Theta(\frac{1}{\sqrt{n}})$, $\mathrm{Inf}_i(parity) = 1$. Using the Fourier transform, we could express the influence function into analytical form.

Lemma 3.

$$\mathrm{Inf}_i(f) = \sum_{S, S \ni i} \widehat{f}(S)^2$$

Proof.

$$\mathrm{Inf}_i(f) = \mathop{\mathrm{E}}_x \left[\frac{1 - f(x)f(x^i)}{2} \right] = \frac{1}{2} - \frac{1}{2} \mathop{\mathrm{E}}_x[f(x)f(x^i)].$$

Using the Fourier transform, we have

$$\begin{aligned}
\mathop{\mathrm{E}}_x[f(x)f(x^i)] &= \mathop{\mathrm{E}}_x \left[\sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x) \sum_{T \subseteq [n]} \widehat{f}(T)\chi_T(x^i) \right] \\
&= \sum_{S,T \subseteq [n]} \widehat{f}(S)\widehat{f}(T) \mathop{\mathrm{E}}_x \left[\chi_S(x)\chi_T(x^i) \right] \\
&= -\sum_{S \ni i} \widehat{f}(S)^2 + \sum_{S \not\ni i} \widehat{f}(S)^2.
\end{aligned}$$

Then we finally get

$$\begin{aligned}
\mathrm{Inf}_i(f) &= \frac{1}{2}[1 + \sum_{S \ni i} \widehat{f}(S)^2 - \sum_{S \not\ni i} \widehat{f}(S)^2] \\
&= \frac{1}{2}[\sum_{S} \widehat{f}(S)^2 + \sum_{S \ni i} \widehat{f}(S)^2 - \sum_{S \not\ni i} \widehat{f}(S)^2] \\
&= \sum_{S, S \ni i} \widehat{f}(S)^2.
\end{aligned}$$

□

Definition 4. Average sensitivity

$$AS(f) = avg\mathrm{Inf}_i(f) = \sum_{S \subseteq [n]} \frac{|S|}{n} \widehat{f}(S)^2.$$

3 Application: Property testing

Now we consider an application of the Fourier transform of boolean functions. The property testing problem is, given a boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we want to test if it has a property P. Our goal is: find an algorithm, given f, the algorithm make q queries of f. If f has property P, the algorithm outputs 'accept' with probability 1(completeness); the algorithm outputs 'reject' with probability $\delta > 0$ if f is ϵ-far from all $g \in P$(soundness). Here we define function f is ϵ-far from g, if

$$dist(f,g) = \frac{|\{x : f(x) \neq g(x)\}|}{2^n} > \epsilon.$$

More specifically, we consider the problem of linearity testing, where $P = \{\chi_S | S \subseteq [n]\}$. The algorithm in [BLR90] for linearity testing is:

1. Generate $x, y \in \{-1,1\}^n$ randomly.
2. If $f(x)f(y)f(xy) = 1$, return 'accept'.
3. Otherwise, return 'reject'.

We have 3 queries in this algorithm. The completeness of the algorithm is obvious: if $f = \chi_S$, then $\chi_S(x)\chi_S(y)\chi_S(xy) = 1$. For the soundness, we have the following claim.

Claim 1. *If f is ϵ-far from all $g \in P$, $\Pr[\text{'}reject\text{'}] \geq \epsilon$ using algorithm in [BLR90] for linearity testing.*

Proof.

$$\begin{aligned}
\Pr[\text{'}reject\text{'}] &= \Pr[f(x)f(y)f(xy) = -1] \\
&= \mathop{\mathrm{E}}_{x,y}\left[\frac{1 - f(x)f(y)f(xy)}{2}\right] \\
&= \frac{1}{2} - \frac{1}{2}\mathop{\mathrm{E}}_{x,y}[f(x)f(y)f(xy)],
\end{aligned}$$

where

$$\begin{aligned}
\mathop{\mathrm{E}}_{x,y}[f(x)f(y)f(xy)] &= \mathop{\mathrm{E}}_{x,y}\left[\sum_S \widehat{f}(S)\chi_S(x)\sum_T \widehat{f}(T)\chi_T(y)\sum_U \widehat{f}(U)\chi_U(xy)\right] \\
&= \mathop{\mathrm{E}}_{x,y}\left[\sum_{S,T,U} \widehat{f}(S)\widehat{f}(T)\widehat{f}(U)\chi_S(x)\chi_T(y)\chi_U(x)\chi_U(y)\right] \\
&= \sum_{S,T,U} \widehat{f}(S)\widehat{f}(T)\widehat{f}(U)\mathop{\mathrm{E}}_{x}[\chi_S(x)\chi_U(x)]\mathop{\mathrm{E}}_{y}[\chi_T(y)\chi_U(y)] = \sum_S \widehat{f}(S)^3.
\end{aligned}$$

If f is ϵ-far from all $g \in P$, assume $\Pr[\text{'}reject'] < \epsilon$, then we have

$$\epsilon > \Pr[\text{'}reject'] = \frac{1}{2} - \frac{1}{2}\sum_S \widehat{f}(S)^3,$$

which indicates

$$\widehat{f}(S)^3 > 1 - 2\epsilon.$$

On the other hand,

$$\sum_S \widehat{f}(S)^3 \leq \sum_S \widehat{f}(S)^2 \times \max_S\{\widehat{f}(S)\} = \max_S\{\widehat{f}(S)\}.$$

Then we know that, $\exists S = S^*$, such that $\widehat{f}(S^*) > 1 - 2\epsilon$. If we consider χ_{S^*},

$$dist(f, S^*) = \mathop{\mathrm{E}}_x\left[\frac{1 - f(x)\chi_{S^*}}{2}\right] = \frac{1}{2} - \frac{1}{2}\langle f, \chi_{S^*}\rangle = \frac{1}{2} - \frac{1}{2}\widehat{f}(S^*) < \epsilon.$$

This contradicts with the f is ϵ-far from all $g \in P$. □

Note that, we could boost the probability to high probability by repeating the algorithm $O(\frac{1}{\epsilon})$ times.

References

[BLR90] Blum M, Luby M, Rubinfeld R. Self-testing/correcting with applications to numerical problems. *Proceedings of the twenty-second annual ACM symposium on Theory of computing (STOC)*, 1990: 73–83.