Analysis of Boolean Functions

Based on chapter 1 of “Analysis of Boolean Functions” by Ryan O’Donnell

Boolean functions as polynomials

- $f: \{0,1\}^n \to \{0,1\}$ or more generally, $f: \{0,1\}^n \to \mathbb{R}$.
- Equivalently, $f: \{-1,1\}^n \to \{-1,1\}$ or more generally, $f: \{-1,1\}^n \to \mathbb{R}$.

- Any such function can be written as a polynomial.
- Example 1: $f: \{-1,1\} \to \mathbb{R}$
- Polynomial $\frac{1}{2}(1-x) = \begin{cases} 0 \; if \; x = 1 \\ 1 \; if \; x = -1 \end{cases}$ and $\frac{1}{2}(1+x) = \begin{cases} 1 \; if \; x = 1 \\ 0 \; if \; x = -1 \end{cases}$
- Therefore, $f(x) = f(1) \cdot \frac{1}{2}(1+x) + f(-1) \cdot \frac{1}{2}(1-x)$

- For any $a \in \{-1,1\}^n$

$$\left(\frac{1+a_1x_1}{2}\right)\cdot\left(\frac{1+a_2x_2}{2}\right)\cdot\cdots\cdot\left(\frac{1+a_nx_n}{2}\right) = \begin{cases} 1\ if\ x_i = a_i\ \forall i \in [n] \\ 0\ otherwise \end{cases}$$

- Therefore,

$$f(x) = \sum_{a\in\{-1,1\}^n} f(a)\Pi_{i\in[n]}\left(\frac{1+a_ix_i}{2}\right)$$

- Multilinear polynomial, i.e., no x_i appears squared, cubed, etc.
- Fourier expansion $f(x) = \sum_{S\subseteq[n]} \hat{f}_S\chi_S$ where $\chi_S \stackrel{\text{def}}{=} \Pi_{i\in S}x_i$.
- This representation is unique (why?)

- $\hat{f}_S$ is the Fourier coefficient of f on S.

Examples

- $\max(x_1, x_2)$ is -1 if both $x_1 = x_2 = -1$, and is 1 otherwise

$$\max(x_1, x_2)$$
$$= \frac{1}{4}(1 + x_1)(1 + x_2) + \frac{1}{4}(1 - x_1)(1 + x_2) + \frac{1}{4}(1 + x_1)(1 - x_2) - \frac{1}{4}(1 - x_1)(1 - x_2)$$
$$= \frac{1}{2} + \frac{1}{2}x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_1x_2$$

- Parity function: $f(x_1, x_2, x_3)$ equal to -1 if x has an odd number -1, and is equal to 1 otherwise.

$$f(x_1, x_2, x_3) = x_1x_2x_3$$

- $f(x_1, \ldots, x_n) = \Pi_{i\in[n]}x_i = \chi_{[n]}$

Parity functions

- The set of all functions $f: \{-1,1\}^n \to \mathbb{R}$ forms a vector space of dimension 2^n.
- For two functions $f, g: \{-1,1\}^n \to \mathbb{R}$, define
$$\langle f, g \rangle \stackrel{\text{def}}{=} \mathrm{E}_{x \sim \{-1,1\}^n}[f(x)g(x)]$$
- Then,
$$\langle \chi_S, \chi_T \rangle = \mathrm{E}_{x \sim \{-1,1\}^n}\left[\Pi_{i \in S} x_i \Pi_{j \in T} x_j\right] = \mathrm{E}_{x \sim \{-1,1\}^n}\left[\Pi_{i \in S \cap T} x_i^2 \Pi_{j \in S \Delta T} x_j\right]$$
$$= \Pi_{j \in S \Delta T} \mathrm{E}_{x_j \sim \{-1,1\}}\left[x_j\right]$$
$$\langle \chi_S, \chi_T \rangle = \begin{cases} 1 \ if \ S = T \\ 0 \ if \ S \neq T \end{cases}$$
- Theorem: $\{\chi_S\}_{S \subseteq [n]}$ forms an orthonormal basis for the vector space of functions $f: \{-1,1\}^n \to \mathbb{R}$.
- $\{\chi_S\}_{S \subseteq [n]}$ are the eigenfunctions of the hypercube graph.

- Proposition: $\hat{f}_S = \langle f, \chi_S \rangle$

$$\langle f, \chi_S \rangle = \left\langle \sum_{T \subseteq [n]} \hat{f}_T \chi_T, \chi_S \right\rangle = \mathrm{E}_x \left[\left(\sum_{T \subseteq [n]} \hat{f}_T \chi_T(x) \right) \cdot \chi_S(x) \right] = \sum_{T \subseteq [n]} \hat{f}_T \langle \chi_T, \chi_S \rangle = \hat{f}_S$$

- $\|f\|_2 = \sqrt{\langle f, f \rangle} = \sqrt{\mathrm{E}_x[f(x)^2]}$
- For $f: \{-1,1\}^n \to \{-1,1\}$, we have $\|f\|_2 = \sqrt{\mathrm{E}_x[f(x)^2]} = 1$

- For functions $f, g: \{-1,1\}^n \to \{-1,1\}$, $\mathrm{dist}(f, g) \stackrel{\text{def}}{=} \Pr_x[f(x) \neq g(x)]$.

Then

$$\langle f, g \rangle = \mathop{\mathrm{E}}_x[f(x)g(x)] = \Pr_x[f(x) = g(x)] - \Pr_x[f(x) \neq g(x)]$$
$$= 1 - 2 \Pr_x[f(x) \neq g(x)] = 1 - 2\mathrm{dist}(f, g)$$

- *Plancherel's Theorem*

$$\langle f, g\rangle = \left\langle \sum_{S\subseteq[n]} \hat{f}_S\chi_S , \sum_{T\subseteq[n]} \hat{g}_T\chi_T \right\rangle = \mathrm{E}_x\left[\sum_{S,T\subseteq[n]} \hat{f}_S\hat{g}_T\chi_S(x)\chi_T(x)\right]$$

$$= \sum_{S,T\subseteq[n]} \hat{f}_S\hat{g}_T\,\mathrm{E}_x[\chi_S(x)\chi_T(x)] = \sum_{S\subseteq[n]} \hat{f}_S\hat{g}_S$$

- *Parseval's Theorem*: $\langle f, f\rangle = \sum_{S\subseteq[n]} \hat{f}_S^2$

- Let $g(x) = 1\forall x$. Then $\hat{g}_S = \begin{cases} 1\ if\ S = \emptyset \\ 0\ otherwise \end{cases}$.

Then

$$\mathrm{E}_x[f(x)] = \langle f, g\rangle = \sum_S \hat{f}_S\hat{g}_S = \hat{f}_\emptyset$$

$$\mathrm{var}(f) = \mathrm{E}_x[f(x)^2] - (\mathrm{E}_x[f(x)])^2 = \langle f, f\rangle - \hat{f}_\emptyset^2 = \sum_{S \neq \emptyset} \hat{f}_S^2$$

$$cov(f, g) = \langle f - \mathrm{E}[f], g - \mathrm{E}[g]\rangle = \mathrm{E}[fg] - \mathrm{E}[f]\,\mathrm{E}[g] = \sum_{S \neq \emptyset} \hat{f}_S \hat{g}_S$$

$$f: \{0,1\}^n \to \mathbb{R}$$

- For $b \in \mathbb{F}_2$, $\chi(b) \stackrel{\text{def}}{=} (-1)^b$, i.e. $\chi\left(0_{\mathbb{F}_2}\right) = 1$ and $\chi\left(1_{\mathbb{F}_2}\right) = -1$
- For $S \subseteq [n]$, $\chi_S \stackrel{\text{def}}{=} \Pi_{i \in S} \chi(x_i) = (-1)^{\sum_{i \in S} x_i}$

- Therefore, $f = \sum_{S \subseteq [n]} \hat{f}_S \chi_s$ (verify)

Property Testing for Boolean Functions

- Given black-box access to a function $f: \{-1,1\}^n \to \{-1,1\}$ test whether it has property P, e.g., f is a *linear function*, etc.
- In many cases, checking whether f has the property P can be done easily in time $2^{O(n)}$.

Goal:

- If f has property P, output Yes with "high" probability.
- If f is "far" from property P, output No with "high" probability.

- $f, g: \{-1,1\}^n \to \{-1,1\}$ are ϵ-close if $\mathrm{dist}(f,g) \le \epsilon$. Let $\wp$ be the set of all functions satisfying property P. Then f is ϵ-close to property P if $\mathrm{dist}(f, \wp) \stackrel{\text{def}}{=} \min_{g \in \wp} \mathrm{dist}(f,g) \le \epsilon$

$f(x) = 1 \; \forall x$

<u>*Algorithm*</u>

- Sample k independent random inputs $x_1, \ldots, x_k$, and output Yes if $f(x_i) = 1 \; \forall i$, else output No.
- If $f = 1$, then $\Pr[Alg\ outputs\ Yes] = 1$.

- If f is ϵ-far from 1, then $\Pr_x[f(x) \neq 1] \geq \epsilon$, i.e., f evaluates to 1 on at most $1 - \epsilon$ fraction of inputs.

$$\Pr[Alg\ outputs\ Yes] \leq (1-\epsilon)^k$$

- For $k \geq 1/\epsilon$, $\Pr[Alg\ outputs\ Yes] \leq 1/e$. Larger value of k can make probability smaller.

Linearity Testing

$g: \mathbb{F}_2^n \to \mathbb{F}_2$ is a linear function if

- $g(x+y) = g(x) + g(y)\ \forall x, y \in \mathbb{F}_2^n$.
- Equivalently, $g(x) = \sum_{i \in [n]} a_i x_i$ for some $a \in \mathbb{F}_2^n$ (verify).

<u>Algorithm</u> (**B**lum, **L**uby, **R**ubinfeld - 1990)

- Sample $x, y \sim \mathbb{F}_2^n$. If $g(x+y) = g(x) + g(y)$ output YES, else output No.
- If g is linear, then $\Pr[BLR\ test\ outputs\ Yes] = 1$.
- Theorem (BLR-90): If test accepts with probability $1 - \epsilon$, then g is ϵ-close to linear.

Testing parity functions

- Equivalent to testing whether $f : \{-1,1\}^n \to \{-1,1\}$ is a parity function (verify).
- If $f = \chi_S$ for some $S \subseteq [n]$, then $f(x \cdot y) = f(x)f(y)\ \forall x, y \in \{-1,1\}^n$ where $(x \cdot y)_i \stackrel{\text{def}}{=} x_i y_i$.

- BLR Test: Sample $x, y \sim \{-1,1\}^n$ and check whether $f(x \cdot y) = f(x)f(y)$.
- Test outputs Yes if $f(x)f(y) = f(x \cdot y)$ or equivalently, $f(x)f(y)f(x \cdot y) = f(x \cdot y)^2 = 1$.
- $\frac{1}{2} + \frac{1}{2}f(x)f(y)f(x \cdot y)$ is an indicator of whether Test outputs Yes.

$$\Pr[Test\ output\ Yes] = E_{x,y}\left[\frac{1}{2} + \frac{1}{2}f(x)f(y)f(x \cdot y)\right]$$

$$\Pr[Test\ outputs\ Yes] = E_{x,y}\left[\frac{1}{2} + \frac{1}{2} f(x) f(y) f(x \cdot y)\right]$$

$$= E_{x,y}\left[\frac{1}{2} + \frac{1}{2}\left(\sum_{S\subseteq[n]} \hat{f}_S \Pi_{i\in S} x_i\right)\left(\sum_{T\subseteq[n]} \hat{f}_T \Pi_{i\in T} y_i\right)\left(\sum_{U\subseteq[n]} \hat{f}_U \Pi_{i\in U} x_i y_i\right)\right]$$

$$= \frac{1}{2} + \frac{1}{2}\sum_{S,T,U\subseteq[n]} \hat{f}_S \hat{f}_T \hat{f}_U\, E_{x,y}[\Pi_{i\in S} x_i \Pi_{i\in T} y_i \Pi_{i\in U} x_i y_i]$$

$$= \frac{1}{2} + \frac{1}{2}\sum_{S,T,U\subseteq[n]} \hat{f}_S \hat{f}_T \hat{f}_U\, E_{x,y}[\underbrace{\Pi_{i\in S\cap U} x_i^2}_{=1} \Pi_{i\in S\Delta U} x_i \underbrace{\Pi_{i\in T\cap U} y_i^2}_{=1} \Pi_{i\in T\Delta U} y_i]$$

$$= \frac{1}{2} + \frac{1}{2}\sum_{S,T,U\subseteq[n]} \hat{f}_S \hat{f}_T \hat{f}_U\, E_x[\Pi_{i\in S\Delta U} x_i]\, E_y[\Pi_{i\in T\Delta U} y_i] = \frac{1}{2} + \frac{1}{2}\sum_{S\subseteq[n]} \hat{f}_S^3$$

- Last step uses $E_x[\Pi_{i\in S\Delta U} x_i] = 0$ if $S \neq U$, and $E_y[\Pi_{i\in T\Delta U} y_i] = 0$ if $T \neq U$.

- Using Parseval's theorem,

$$\Pr[Yes] = \frac{1}{2} + \frac{1}{2}\sum_{S\subseteq[n]} \hat{f}_S^3 \leq \frac{1}{2} + \frac{1}{2}\left(\sum_{S\subseteq[n]} \hat{f}_S^2\right) \cdot \max_{S\subseteq[n]} \hat{f}_S = \frac{1}{2} + \frac{1}{2}\max_{S\subseteq[n]} \hat{f}_S$$

- Therefore, if $\Pr[Yes] \geq 1-\epsilon$, then $\max_{S\subseteq[n]} \hat{f}_S \geq 1 - 2\epsilon$.
- Recall that $\hat{f}_S = \langle f, \chi_S\rangle = 1 - 2\,\mathrm{dist}(f, \chi_S)$.
- Therefore, for $S^* \stackrel{\text{def}}{=} \operatorname{argmax}_{S\subseteq[n]} \hat{f}_S$, we have $\mathrm{dist}(f, \chi_{S^*}) \leq \epsilon$

- Computing $\chi_{S^*}(x)$: We have $f(x) = \chi_{S^*}(x)$ for $(1-\epsilon)$-fraction of inputs. What about remaining ϵ-fraction of inputs?

Local correctability

Theorem: Suppose f is ϵ-close to χ_{S^*}. For every $x \in \{-1,1\}^n$, following algorithm outputs $\chi_{S^*}(x)$ with probability at least $1 - 2\epsilon$.

Algorithm(x)

1. Sample $y \sim \{-1,1\}^n$.
2. Output $f(y)f(x \cdot y)$

- y and $y \cdot x$ are uniformly distributed in $\{-1,1\}^n$ (but are not independent).
- $\Pr_y[f(y) = \chi_{S^*}(y)] \geq 1 - \epsilon$ and $\Pr_y[f(y \cdot x) = \chi_{S^*}(y \cdot x)] \geq 1 - \epsilon$
- Therefore, with probability at least $1 - 2\epsilon$ (over the choice of y), we have $f(y) = \chi_{S^*}(y)$ and $f(y \cdot x) = \chi_{S^*}(y \cdot x)$. In such an event,
$$f(y)f(y \cdot x) = \chi_{S^*}(y)\chi_{S^*}(y \cdot x) = \chi_{S^*}(x)$$

Analysis of Boolean Functions

- Social Choice theory
- Learning Theory
- Hardness of approximation, PCPs
- Many more ..

- See “Analysis of Boolean Functions” by Ryan O’Donnell