Notes

Debarshi Chanda

July 22, 2024

Chapter 1

Lecture 1: Introduction

Relveant resource: [Wol08],[OD14].

1.1 Scope of The Course

Two possible formats to represent boolean functions:

- $f : \mathbf{F}_2^n \to \mathbf{F}_2$ / $f : \{0,1\}^n \to \{0,1\}$
- $f : \{-1,1\}^n \to \{-1,1\}$

Topics to be covered:

- Influence
- Noise Sensitivity
- Polynomial approximation
- Hypercontractivity
- Invariance Principal

Here, we list some of the interesting problems that we would deal with in this course:

Interesting Problem in Additive Number Theory

Problem: Consider $S \subset \mathbb{N}$, we define density of the set as $density(S) = \frac{|S|}{N}$. The question we study here is at what density for the set, it will always have infinitely many arithmetic progression, say of length 3.

Results: The Density of primes is known to be $O(\frac{1}{\log N})$. Green-Tao theorem states that the set of primes have arbitrary arithmetic progressions of length $\geq k$, for all $k \in \mathbb{N}$. Recent result of [KM24] showed that an arithmetic progression exists even when density is $<< \frac{1}{\log N}$. The best known lower bound shows that there exists set S with density $\frac{1}{e^{c\sqrt{\log n}}}$ such that it does not contain infinitely many sequences of length 3. An interesting exposition: [BS23].

Social Choice Theory - Arrows Impossibility Result

Problem: Social choice theory studies the mechanisms for collective decision-making and their impacts. One important question in this domain is to determine whether certain voting mechanisms result in logically coherent outcomes.

Results: Kenneth Arrow in [Arr50] proposed an axiomatic framework that generalizes all ranked-choice voting mechanisms and showed that no ranked-choice voting can produce logically coherent results.

Goldreich Levin Theorem

1.2 Introduction

We will interchangeably use the two notations of boolean functions. For now, let us consider the family of maps $\mathcal{F} = \{f | f : \{-1,1\}^n \to \{-1,1\}\}$.

Lemma 1. *Every function $f \in \mathcal{F}$ can be expressed as $f(x) = \sum_{a \in \{-1,1\}^n} a(x)f(a)$.*

Proof. We define $a(x)$ to be an indicator function defined as $a(x) = \begin{cases} 1 & if x = a \\ 0 & otherwise \end{cases}$. Now, for all $x \in \{-1,1\}^n$, we have:

$$f(x) = \sum_{a \in \{-1,1\}^n} a(x)f(a) = f(x)$$

□

Observe that this proof can be extended to the set of real-valued boolean functions, i.e. this representation holds for all functions in $\mathcal{F}_{\mathbb{R}} = \{f | f : \{-1, 1\}^n \to \mathbb{R}\}$. The indicator functions $a(x)$ can be defined as $a(x) = \left(\frac{a_1 x_1 + 1}{2}\right)\left(\frac{a_2 x_2 + 1}{2}\right)\left(\frac{a_3 x_3 + 1}{2}\right) \ldots$.

Lemma 2. *Every function $f \in \mathcal{F}$ can be written as $f = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S$.*

Proof. Observe that each $a(x) = \Pi_{i \in [n]} \frac{1 + a_i x_i}{2}$. This product is actually a linear combination of monomials of the form $\chi^S = \Pi_{i \in S} x_i$ where $S \subseteq [n]$. Thus, through this representations of $a(x)$, we can also write f as $f(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi^S$. □

Again, this statement holds for all functions in the set $\mathcal{F}_R$. The following lemma states that this set is a vector space.

Lemma 3. *$\mathcal{F}_{\mathbb{R}}$ forms a vector space of dimension 2^n.*

Proof Idea. Let us recall that any function $f \in \mathcal{F}_{\mathbb{R}}$ can be expressed as $f(x) = \sum_{a \in \{-1,1\}^n} a(x)$ $f(a)$. Equipped with this, we can see that pointwise addition and scalar multiplication makes $\mathcal{F}_{\mathbb{R}}$ a vector space.

Also note that, the indicator functions $a(x)$ spans the space and are linearly independent (*Two indicator functions can not combine to provide a new indicator function*). Thus, $\{a(x) | a \in \{-1, 1\}^n\}$ forms a basis of the vector space $\mathcal{F}_{\mathbb{R}}$, making its dimension 2^n. □

Lemma 4. *The monomials $\{\chi_S\}_{S \in [n]}$ forms a basis of sF.*

Proof. By lemma 2,we know that the set of monomials $\{\chi_S | S \subseteq [n]\}$ spans $\mathcal{F}_{\mathbb{R}}$. Additionally, there are 2^n such monomials, matching the dimension of the vector space $\mathcal{F}_{\mathbb{R}}$. Hence, the set of monomials $\{\chi_S\}_{S \in [n]}$ forms a basis for $\mathcal{F}_{\mathbb{R}}$. □

Fourier Expansion Theorem

We have seen that any $f \in \mathcal{F}$ can be written as:

$$f = \sum_{s \subseteq [n]} \hat{f}(s)\chi_S$$

Here $\hat{f}(s)$ is known as *Fourier Coefficient* and the monomials χ_S are known as the *Characteristic Functions* or *Characters*. By lemma 4, this representation is unique.

The domain we will be working with is an n-dimensional hamming cube. We can define a probability simplex over it. There are two possible ways to define two equivalent distributions over the domain:

- An Uniform distribution over the vectors $\{-1, 1\}^n$.

- Each coordinate x_i of x is fixed independently to -1 or 1 with equal probability.

Definition 5 (Expectation over $\mathcal{F}$)**.**

$$\mathbb{E}[f] = \frac{\sum_{a \in \{-1,+1\}^n} f(a)}{2^n}$$

Definition 6 (Inner Product)**.** The inner product for two functions $f, g \in \mathcal{F}$ is defined as $\langle f \,,\, g \rangle = \mathbb{E}[f(x)g(x)]$.

Definition 7 (p-th norm)**.** The p-th norm of a function $f \in \mathcal{F}$ is defined as $\|f\|_p = (\mathbb{E}[f^p(x)])^{\frac{1}{p}}$.

Spectral Norm: When $p = 1$, then the norm $\|f\|_1$ is called the spectral norm of f.

Lemma 8. *Calculate* $\mathbb{E}[\chi_S] = \begin{cases} 1 & if S = \emptyset \\ 0 & otherwise \end{cases}$

Proof. if $\chi_S = \emptyset$, $\mathbb{E}[\chi_S] = E[1] = 1$. When $S \neq \emptyset$,

$$\begin{aligned} \mathbb{E}[\chi_S] &= \mathbb{E}[\Pi_{i \in S} x_i] \\ &= \Pi_{i \in S} E[x_i] \\ &= 0 \end{aligned}$$

Because, if $S \neq \emptyset$, then we use the independence of x_i's to break down the product int terms of $\mathbb{E}[x_i]$, each of which has value 0. □

Theorem 9. *The characteristic functions form an orthonormal basis.*

Proof.

$$\begin{aligned} \chi_S(x)\chi_T(x) &= \Pi_{i \in S} x_i \Pi_{i \in T} x_i \\ &= \Pi_{i \in S \cap T} x_i^2 \Pi_{i \in S \Delta T} x_i \\ &= \Pi_{i \in S \Delta T} x_i \\ &= \chi_{X \Delta Y} \end{aligned}$$

Now, we have:

$$\begin{aligned}\langle \chi_S \, , \, \chi_T \rangle &= \mathbb{E}[\chi_S \chi_T] \\ &= \mathbb{E}[\chi_{S\Delta T}] \\ &= \begin{cases} 0 & \text{when } S \neq T \\ 1 & \text{when } S = T \end{cases}\end{aligned}$$

□

Lemma 10 (Fourier Coefficient Formula)**.** $\hat{f}(S) = \langle f \, , \, \chi_S \rangle$

Proof.

$$\begin{aligned}\langle f \, , \, \chi_S \rangle &= \left\langle \sum_{T \subseteq [n]} \hat{f}(T) \chi_T \, , \, \chi_S \right\rangle \\ &= \sum_{T \subseteq [n]} \hat{f}(T) \left\langle \chi_T \, , \, \chi_S \right\rangle \\ &= \hat{f}(S)\end{aligned}$$

The last equality is due to the fact that $\langle \chi_T \, , \, \chi_S \rangle = 1$ only when $S = T$. □

Lemma 11. *For $f, g \in \mathcal{F}$, we have:*

$$\begin{aligned}\langle f \, , \, g \rangle &= \sum_{S \subseteq [n]} \hat{f}(S) \hat{g}(S) \\ \|f\|_2 &= \sum_{S \subseteq [n]} (\hat{f}(S))^2\end{aligned}$$

Proof.

$$\begin{aligned}\langle f \, , \, g \rangle &= \mathbb{E}[f(x) g(x)] \\ &= \mathbb{E}[\sum_{S \subseteq [n]} \hat{f}(S) \chi_S \cdot \sum_{T \subseteq [n]} \hat{g}(T) \chi_T] \\ &= \mathbb{E}[\sum_{S \subseteq [n]} \hat{f}(S) \hat{g}(S)] \\ &= \sum_{S \subseteq [n]} \hat{f}(S) \hat{g}(S)\end{aligned}$$

□

Definition 12 (distance(f,g))**.** For $f, g \in \mathcal{F}$, we define their distance as:

$$distance(f, g) = \frac{|\{a \in \{-1,1\}^n | f(a) \neq g(a)\}|}{2^n} = \mathbb{P}(f(x) \neq g(x))$$

Lemma 13. $\langle f \ , \ g \rangle = \mathbb{P}[f(x) = g(x)] - \mathbb{P}[f(x) \neq g(x)] = 1 - 2 \times distance(f, g)$

Proof. This comes from the simple observation that $\langle f \ , \ g \rangle = \mathbb{E}[f(x)g(x)]$ and $f(x)g(x)$ is 1 if $f(x) = g(x)$ and -1 otherwise. □

Lemma 14. $\mathbb{E}[f] = \hat{f}(\phi)$

Proof.

$$\begin{aligned}\mathbb{E}[f] &= \mathbb{E}[f, \mathbf{1}] \\ &= \left\langle f \ , \ \chi_\phi \right\rangle \\ &= \hat{f}(\phi)\end{aligned}$$

□

Lemma 15. $\mathrm{Var}[f] = \sum_{S \subseteq [n], s \neq \phi} \hat{f}(S)^2$

$$\begin{aligned}\mathrm{Var}[f] &= \mathbb{E}[f^2] - \mathbb{E}[f]^2 \\ &= \sum_{S \subseteq [n]} \hat{f}^2(s) - \hat{f}(\phi)^2 \\ &= \sum_{S \subseteq [n], S \neq \phi} \hat{f}(S)^2\end{aligned}$$

Chapter 2

Lecture 2: Introduction

Recall the following facts that will be relevant to our class today:

- Every real-valued boolean function $f : \{-1,1\}^{\mathbb{R}} \to \mathbb{R}$ can be written as a linear combination of monomials of the form $\{X^S\}_{S \subseteq [n]}$.
- $\{X^S\}_{S \subseteq [n]}$ forms an orthonormal basis for $f : \{-1,1\}^{\mathbb{R}} \to \mathbb{R}$.
- $X^S X^T = X^{S \Delta T}$

Let us now consider the setup of $f : \{-1,1\}^n \to \{-1,1\}$ to the setup of $f : \{0,1\}^n \to \{0,1\}$. We consider the functions $\mathcal{F}_R = \{f | f : \{-1,1\}^n \to \{-1,1\}\}$ and $\mathcal{F}_F = \{f | f : \{0,1\}^n \to \{0,1\}\}$. We consider the following bijection between $f \in \mathcal{F}_R$ and $\hat{f} \in \mathcal{F}_F$ as:

$$\hat{f}(x_1, x_2, ..., x_n) = f((-1)^{x_1}, (-1)^{x_2}, ..., (-1)^{x_n}).$$

Now from a fourier analysis on this transofrmed function yields:

$$\begin{aligned}
&f((-1)^{x_1}, (-1)^{x_2}, ..., (-1)^{x_n}) \\
&= \sum_{S \subseteq [n]} \hat{f}(S) \Pi_{i \in S} (-1)^{x_i} \\
&= \sum_{S \subseteq [n]} \hat{f}(S) (-1)^{\sum_{i \in S} x_i}
\end{aligned}$$

We can also represent the set S as an indicator vector $\alpha_S \in \mathbb{F}_2^n$. We remove the subscript S when it is clear from the context. Then, we can state the result above as:

$$\hat{f} = \sum_{\alpha \in \mathbb{F}_n^2} \hat{f}(\alpha) \chi_\alpha$$

Where $\hat{f}(\alpha) = \hat{f}(S)$ and $\chi_\alpha(x) = (-1)^{\langle \alpha\,, x \rangle}$. We now have the following two results:

- $\alpha + \beta$ in this setup is equivalent to the operation $S_\alpha \Delta S_\beta$.
- $\chi_\alpha(x + y) = \chi_\alpha(x)\chi_\alpha(y)$
- $\chi_{\alpha+\beta} = \chi_\alpha(x)\chi_\beta(y)$
- The above two facts combined ensures that $(\chi_\alpha, \cdot)$ is a group and it is isomorphic to $\mathbb{F}_n^2$.

Theorem 16. *The group $\hat{\mathbb{F}}_2^n = (\chi_\alpha, \cdot)$ is isomorphic to the group $\mathbb{F}_2^n$. This is known as the dual group of $\mathbb{F}_2^n$.*

Let us now begin the study of testing properties of the functions of the form $f : \mathbb{F}_n^2 \to \mathbb{F}$ under the query model where we have black-box query access to the functions.

Definition 17 (Property $\mathcal{P}$)**.** We define a function $f \in \mathcal{F}_\mathbb{F}$ to have a property $\mathcal{P}$ if $f \in \mathcal{P} \subseteq \mathcal{F}_\mathbb{F}$.

Definition 18 (Distance to a property $\mathcal{P}$)**.** We define distance to a property w.r.t. a distance $d : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$as:

$$dist(f, \mathcal{P}) = \min_{f \in \mathcal{F}} d(g, f)$$

For most of this course, we will be considering the distance between functions as defined in the last lecture to be $d(f, g) = \mathbb{P}_{x \in \mathbb{F}_2^n} f(x) \neq g(x) = \sum_{x \in \mathbb{F}_2^n} \|f(x) = g(x)\|^2 / (4 \times 2^n)$. For completeness note that:

$$\begin{aligned} &d(f, g) \\ =&\mathbb{P}_{x \in \mathbb{F}_2^n} f(x) \neq g(x) \\ =&\frac{\sum_{x \in \mathbb{F}_2^n} \|f(x) = g(x)\|^2}{4 \times 2^n} \\ =&\frac{\sum_{x \in \mathbb{F}_2^n} \|f(x) = g(x)\|}{2 \times 2^n} \end{aligned}$$

Let us consider one of the simplest property $\mathcal{P}$ to be the set of constant functions. In this case, we have to actually check exactly for this property, then we have to query 2^n points as it can deviate at a single point. However, can we do better if we have some **promise** on the input, say of the form: *function f is either constant or ϵ-far from constant*? Note that **ϵ-far** from a property p means that a function f is at a distance at least ϵ from a property p, i.e. $dist(f, \mathcal{P}) \geq \epsilon$.

A possible idea to test for constant function under the stated promise can be to pick two points $x, y \in \mathbb{F}_2^n$ uniformly and independently at random, we check if they are same or different.

Algorithm 1 Atomic test for constant functions

Pick $x, y \in \mathbb{F}_2^n$ uniformly and independently at random
If $f(x) = f(y)$, return 1. Otherwise, return 0.

Lemma 19. *If a function $f \in \mathcal{F}$ is ϵ-far from being a constant function, then the test 1 returns 0 with probability at least $<? >$.*

Algorithm 2 Test for constant functions

?

Theorem 20. *If a function $f \in \mathcal{F}$ is ϵ-far from being a constant function, then the test 2 returns 0 with probability at least $<? >$.*

Now, let us consider the property $\mathcal{P}$ that is the set of all linear functions in $\mathcal{F}$, i.e. $\mathcal{P} = \{f \in \mathcal{F} | f(x+y) = f(x) + f(y)\}$. Equivalently, we can define it as $\mathcal{P} = \{f \in \mathcal{F} | \exists \alpha \in \mathbb{F}_2^n \text{ s.t. } f(x) = \langle \alpha \,, x \rangle\}$. We denote these functions l_α.

Lemma 21. $\{f \in \mathcal{F} | \exists \alpha \in \mathbb{F}_2^n \text{ } s.t. \text{ } f(x) = \langle \alpha \,, x \rangle\} = \{f \in \mathcal{F} | f(x+y) = f(x) + f(y)\}$

Proof Sketch. $\alpha_i = f(e_i)$ □

For property testing, we require more robust definitions to account for the cases of ϵ-close and ϵ-far. Let us start with the notion of distance d directly.

Definition 22 (ϵ-close to linear)**.** We say f *is ϵ-close to linear* if $\exists l_\alpha$ such that $\mathbb{P}_{x \in \mathbb{F}_2^n}[l_\alpha(x) \neq f(x)] \leq \epsilon$ or equivalently, $\mathbb{P}_{x \in \mathbb{F}_2^n}[l_\alpha(x) = f(x)] \geq \epsilon$.

Now, we look for an equivalent definition where a function f is ϵ-close to linear, we will have $\mathbb{P}_{x,y \sim F_2^n}[f(x) + f(y) = f(x+y)] \geq 1 - \epsilon$. If we have this definition to be equivalent to the earlier robust definition 22, then the linearity property will be testable (with $poly(\frac{1}{\epsilon})$ samples). Consider the following algorithm proposed in ¡cite¿.

Algorithm 3 Atomic Tester for Linearity—BLR Test

Require: Black-box access to $f : \mathbb{F}_2^n \to \{-1, 1\}$
Pick $x, y \in \mathbb{F}_2^n$ uniformly and independently at random
If $f(x+y) = f(x)f(y)$, return 1; otherwise return 0

Theorem 23. *If f is ϵ-far from linear, then with probability $\Omega(\epsilon)$, BLR test detects it.*

Theorem 24. *If f is ϵ-close to linear as per definition 22, then $\mathbb{P}_{x,y \sim F_2^n}[f(x) + f(y) = f(x+y)] \geq 1 - \epsilon$.*

Proof. Recall the fact that $\langle f \,, g \rangle = 1 - 2d(f, g)$, and $\hat{f}(\alpha) = \langle f \,, \chi_\alpha \rangle \rangle$. Now, observe that

$\forall \alpha \in \mathbb{F}_2^n$, we have $\hat{f}(\alpha) = \langle f \, , \, \chi_\alpha \rangle = 1 - 2d(f, \chi_\alpha)$. Now observe that χ_α are essentially all the linear functions in $\mathcal{F}$. Thus, we have:

$$\max_\alpha \hat{f}(\alpha) \leq 1 - 2\epsilon \tag{2.1}$$

We also have,

$$\begin{aligned}
&\mathbb{E}_{x,y}[f(x)f(y)f(x+Y)] \\
=&\mathbb{P}[f \text{passes BLR}] - \mathbb{P}[f \text{fails BLR}] \\
=&1 - 2\mathbb{P}[f \text{fails BLR}]
\end{aligned}$$

We now use the fourier expansion on the term $f(x)f(y)f(x+y)$,

$$\begin{aligned}
&f(x)f(y)f(x+y) \\
=&\sum_{\alpha,\beta,\gamma} \hat{f}(\alpha)\hat{f}(\beta)\hat{f}(\gamma)\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(x+y)
\end{aligned}$$

Then, we have:

$$\begin{aligned}
&\mathbb{E}_{x,y}[f(x)f(y)f(x+y)] \\
&= \mathbb{E}_{x,y}[\sum_{\alpha,\beta,\gamma} \hat{f}(\alpha)\hat{f}(\beta)\hat{f}(\gamma)\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(x+y)] \\
&= \sum_{\alpha,\beta,\gamma} \hat{f}(\alpha)\hat{f}(\beta)\hat{f}(\gamma)\mathbb{E}_{x,y}[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(x+y)] \\
&= \sum_{\alpha,\beta,\gamma} \hat{f}(\alpha)\hat{f}(\beta)\hat{f}(\gamma)\mathbb{E}_x[\chi_\alpha(x)\chi_\gamma(x)\mathbb{E}_y[\chi_{\beta+\gamma}(y)] \qquad \text{By Linearity of } \chi
\end{aligned}$$

Note that the term $\mathbb{E}_y[\chi_{\beta+\gamma}(y)]$ becomes 0 unless $\beta = \gamma$. Similarly, we can see that $\alpha = \beta$ is necessary for the summation term to be non-zero. Then, the sum can be rewritten as:

$$\begin{aligned}
&\mathbb{E}_{x,y}[f(x)f(y)f(x+y)] \\
&= \mathbb{E}_{x,y}[\sum_{\alpha,\beta,\gamma} \hat{f}(\alpha)\hat{f}(\beta)\hat{f}(\gamma)\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(x+y)] \\
&= \sum_{\alpha,\beta,\gamma} \hat{f}(\alpha)\hat{f}(\beta)\hat{f}(\gamma)\mathbb{E}_{x,y}[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(x+y)] \\
&= \sum_{\alpha\in\mathbb{F}_2^n} \hat{f}(\alpha)^3 \\
&\leq \left(\max_\alpha \hat{f}(\alpha)\right) \sum_\alpha \hat{f}(\alpha)^2 \\
&\leq \max_\alpha \hat{f}^3(\alpha) && \text{By Placherel's Lemma} \\
&\leq 1 - 2\epsilon && \text{By equation 2.1}
\end{aligned}$$

□

Problem 1: Suppose $f \in \mathcal{F}$ is close to some unknown parity function χ_α. Can we find out χ_α?

Sol: Output $x, y \sim \mathbb{F}_2^n$, output $\chi_\alpha(x) = f(y)f(y+x)$.

Chapter 3

Lecture 3: Introduction

3.1 Recap/Questions

In the last lecture, we covered the linearity tests using the following two equivalent definitions:

- $\exists \alpha \in \mathbb{F}_n^2$ such that $\mathbb{P}(f(x) = \langle \alpha\, , \, x \rangle) \geq 1 - \epsilon$
- $\mathbb{P}(f(x) + f(y) = f(x+y) \geq 1 - \epsilon$

Recall that the characters χ_α are the online linear maps in $\mathcal{F}$, with $\chi_{\alpha+\beta} = \chi_\alpha \cdot \chi_\beta$. Note that $f(x) = \chi_\alpha(x) \cdot -1$ are affine maps, not linear. Recall that BLR test only fails when the map is not linear, i.e. has one sided error. Hence, it is amenable to boosting.

Recall the problem discussed in the last class:

Problem 1: Suppose $f \in \mathcal{F}$ is close to some unknown parity function χ_α. Can we find out χ_α where its close in term of l_∞ norm?

Solution: Report $\widehat{\chi_\alpha}(x) = \mathbb{E}_{y \in \mathbb{F}_2^n}[f(y+x)f(y)]$.

Observe that by the definition of ϵ-close, we have:

$$\mathbb{P}_{y \in \mathbb{F}_2^n}[f(y) \neq \chi_\alpha(y)] \leq \epsilon \tag{3.1}$$
$$\mathbb{P}_{y \in \mathbb{F}_2^n}[f(x+y) \neq \chi_\alpha(x+y)] \leq \epsilon \tag{3.2}$$

Combining through union bound, we get:

$$\mathbb{P} y \in \mathbb{F}_2^n [f(y)f(y+x) = \chi_\alpha(x)] \geq 1 - \epsilon$$

Problem 2: Can we estimate the distance of a function f from the nearest linear function?

Idea: If $dist(f, l_\alpha) = \epsilon$, then BLR 'fails' with probability $\geq \epsilon$. Can we do a coin-toss like argument on the BLR test to obtain an estimate of ϵ?

Theorem 25. *Let P_i be the failure probability of BLR test where $\epsilon = dist(f, l_\alpha)$. Then, $P_1 \in [\epsilon, 5\epsilon]$.*

Proof. □

Unique Decoding Theorem
To be added.

Problem 3: Let $f : \mathbb{F}_2^n \to \{-1, 1\}$. Show that the number of characters χ_α such that $\left\|\hat{f}(\chi_\alpha)\right\| > \frac{1}{2}$ is at most one.

3.2 Affine functions and Beyond

Up till now, we have concerned ourselves with checking and approximating linear functions. Can we extend these results to the case of affine functions?

Definition 26 (Affine Functions)**.** There can be multiple alternative definitions for affine functions:

- A function $f \in \mathcal{F}$ is said to be an affine function if it is of the form $f(x) = a + \langle \alpha\, , \, x \rangle$.
- A function $f \in \mathcal{F}$ is said to be an affine function if $\forall x, y, z \in \mathbb{F}_2^n$, we have $f(x+y+z) = f(x) + f(y) + f(z)$

To argue something along the lines of the BLR test, we will require some robust definitions for affine functions as well.

Definition 27 (Robust(ϵ-far) affine functions)**.** Again, we introduce multiple robust definitions for affine functions:

Algorithm 4 Atomic Tester for Affine—Modified BLR Test

Require: Black-box access to $f : \mathbb{F}_2^n \to \{-1, 1\}$
  Pick $x, y, z \in \mathbb{F}_2^n$ uniformly and independently at random
  If $f(x + y + z) = f(x)f(y)f(z)$, return 1; otherwise return 0

Theorem 28. *The modified BLR test (Algorithm 4) returns 0 with probability $q_\epsilon \in [\epsilon, 6\epsilon]$ if $dist(f, l_\alpha) = \epsilon$*

Proof Sketch. Let us again consider the functions to be $f : \mathbb{F}_2^n \to \{-1, 1\}$. Let us consider the event $E = \{f(x)f(y)f(z) = f(x + y + z)\}$. We also have:

$$\mathbf{1}_E(x, y, z) = \frac{1}{2} + \frac{1}{2}f(x)f(y)f(z)f(x + y + z)$$

Then, we have:

$$\begin{aligned}
\mathbb{P}[E] &= \mathbb{E}[\mathbf{1}_E] \\
&= \mathbb{E}[\frac{1}{2} + \frac{1}{2}f(x)f(y)f(z)f(x + y + z)] \\
&= \frac{1}{2} + \frac{1}{2}\mathbb{E}[f(x)f(y)f(z)f(x + y + z)] \\
&= \ldots \\
&= \frac{1}{2} + \frac{1}{2}\sum_{\alpha \in \mathbb{F}_2^n} |\hat{f}^4(\chi_\alpha)|
\end{aligned}$$

Observation: The only possible affine functions are χ_α and $-1 \cdot \chi_\alpha$. $\square$

Roth's Theorem

To be added. Has similar analysis. Doesn't work for four term-arithmetic progressions, for similar reasons why we can't go to qudratic.

However, can we go one step ahead and go to higher degrees, starting with quadratic functions? Turns out, no. Let us consider a particular function $g : \mathbb{F}_n^2 \to \mathbb{F}_2$ defined as:

$$g(x_1, \ldots, x_n) = x_1x_2 + x_3x_4 + \ldots + x_{n-1}x_n$$

And correspondingly the function $G : \mathbb{F}_2^n \to \{-1, 1\}$ defined as:

$$G(x) = (-1)^g(x)$$

We can show that $\forall \alpha \in \mathbb{F}_2^n$,

$$|\hat{G}(\chi_\alpha) = 2^{-\frac{n}{2}}$$

Higher Order Fourier

For the cases of higher degree, similarly for four term extension to the Roth's Theorem, the framework of Higher Order Fourier expansion was introduced by Timothy Gowers.

3.3 Convolutions

Definition 29. Let us consider two functions $f, g \in \mathcal{F}$ with range $\{-1, 1\}$, we define their convolution to be:

$$f * g(x) = \mathbb{E}_{y \in \mathbb{F}_2^n} f(y) g(x - y)$$

Let us list down some important properties of the convolution operation:

- $f * g = g * f$
- $f * g = \langle f \, , \, g \rangle$
- $\widehat{f * g}(\chi_\alpha) = \hat{f}(\chi_\alpha) \cdot \hat{g}(\chi_\alpha)$

Definition 30 (Probability Density Function)**.** $\phi : \mathbb{F}_2^n \to \mathbb{R}_{\geq 0}$ is a probability density function if $\mathbb{E}_{x \sim \mathbb{F}_2^n}[\phi(x)] = 1$.

Probability mass at x w.r.t. ϕ is $\frac{\phi}{2^n}$.

- If x, y are random vectors in $\mathbb{F}_2^n$ drawn according to probability density functions ϕ, ψ, then $X + Y \sim \phi * \psi$.

Chapter 4

Lecture 4

Recall the two formulations $f : \{-1, 1\}^n \to \{-1, 1\}$ and $f : \{0, 1\}^n \to \{-1, 1\}$.

We focus on understanding the functions that captures different voting rules. Let us start with a few examples:

- $Maj_n : \{-1, 1\}^n \to \{-1, 1\}$, defined as $Maj_n(x_1, x_2, ..., x_n) = sign(\sum_{i\in[n]} x_i$
- $AND : \{-1, 1\}^n \to \{-1, 1\}$, defined as:

$$AND(x_1, x_2, ..., x_n) = \begin{cases} -1 & if x = (-1, ..., -1) \\ 1 & otherwise \end{cases}$$

- $OR : \{-1, 1\}^n \to \{-1, 1\}$, defined as:

$$OR(x_1, x_2, ..., x_n) = \begin{cases} +1 & if x = (+1, ..., +1) \\ -1 & otherwise \end{cases}$$

- $Dic_i : \{-1, 1\}^n \to \{-1, 1\}$, defined as $Dic_i(x) = x_i$. These are essentially the characteristic functions of singleton sets. Henceforward referred to as χ_i
- $k - junta : \{-1, 1\}^n \to \{-1, 1\}$, is a k-junta function if $\exists S \subseteq [n]$ with $|S| = k$, and a corresponding $g : \{-1, 1\}^k \to \{-1, 1\}$ such that $\forall x \in \{-1, 1\}^n$, we have $f(x) = g(x_S)$.
- $Threshold_a : \{-1, 1\}^n \to \{-1, 1\}$, given an $a \in \mathbb{R}^n$, is defined as $Threshold_a(x) = sign(\langle a\, ,\, x\rangle)$.

  Pictorially, we can think about this function as a hyperplane separating the boolean hypercube into two sets. Any such labelling that divides the hypercube into two disjoint compact convex sets can be represented through this function, using *Hyperplane Separation Theorem.*

Note: Sometimes called LTF(Linear Threshold Function), with general threshold functions defined w.r.t. function based threshold.

- $Tribes : \{-1, 1\}^s w \to \{-1, 1\}$ is defined to be a (s, w)-tribe function such that it can be represented as the OR of AND of s groups of size w. The function outputs true when any of the group has all 1s.

  Observe that the probability of success, i.e. getting -1 to be equal to getting a probability of failure, i.e. getting 1, we must have $w = \log s$.

Testability

Let us recall the property testing where given a property $\mathcal{P}$ and query access to function f, we check whether $f \in \mathcal{P}$ or $dist(f, \mathcal{P}) > \epsilon$. A property is said to be testable if this can be done with query complexity independent of the size of the truth table, i.e. n.

Conjecture: Characterization of Testability of Boolean Function

Recall that testability in the case of Graph Property Testing is completely characterized by Regularity Lemma. Two separate groups [Eric Blaise and Sourav-Fischer] conjectured that the testability of properties of Boolean Functions can be completely characterized by partial symmetry.

Definition 31 (Symmetric Function)**.** A function $f : \{-1, 1\}^n \to \{-1, 1\}$ is said to be a symmetric function if $f(x) = f(\Pi(x))$, for all permutations Π and $x \in \{-1, 1\}^n$.

Symmetric Function is Testable

Symmetric function is testable. The idea is to choose a random string and check whether the string and a random permutation of itself gives the same output.
Observe that a randomly chosen boolean string of length n has level $n/2$ on average and in $[n/2 - \sqrt{n}, n/2 + \sqrt{n}]$, with high probability. Also note that any permutation functions Π preserves the level of a string.
Exercise: Formalize the algorithm and the proof.

Let us define the obvious partial order on $\{-1, 1\}^n$. For any $x, y \in \{-1, 1\}$, $x \preceq y$ if $\forall i$, $x_i \leq y_i$

Definition 32 (Monotone Function)**.** We define a function $f : \{-1, 1\}^n \to \{-1, 1\}$ to be monotone if $\forall x \preceq y$, $f(x) \leq f(y)$.

Definition 33 (Transitive Symmetric Function)**.** Given a function $f : \{-1, 1\}^n \to \{-1, 1\}$, define $Aut(f) = \{\Pi \in S_n | \forall x, f(x) = f(\Pi(x))\}$. We then define f to be a transitive symmetric function if $\forall i \neq j$, $\exists \Pi \in Aut(f)$ such that $\Pi(i) = j$.

Definition 34 (Unanimous function)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be unanimous if $\forall r \in \mathbb{R}$, we have $f(r, r, ..., r) = r$.

Definition 35 (Good Voting Rule)**.** A voting rule is said to be *"good"* must necessarily satisfy:

- The function should be *monotone*. If more votes go towards one value, the vote should not conclude in the opposite direction.
- The function should be *symmetric*; alternatively a weaker condition can be for the function to be *Transitive symmetric*.
- The function should be *Odd*.
- The function should be *unanimous*.

Exercise: Check whether functions defined at the beginning of this lecture are *good voting rules*.

Chapter 5

Lecture 5

Recall the definition of the linear threshold function:

Definition 36 (Linear Threshold Function)**.** $Threshold_a : \{-1, 1\}^n \rightarrow \{-1, 1\}$, given an $a \in \mathbb{R}^n$, is defined as $Threshold_a(x) = sign(\langle a\, ,\, x\rangle)$.

Similarly, we can define an affine threshold function and general threshold function.

Definition 37 (Affine Threshold Function)**.** $Threshold_{(a,b} : \{-1, 1\}^n \rightarrow \{-1, 1\}$, given an $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, is defined as $Threshold_{(a,b)}(x) = sign(\langle a\, ,\, x\rangle + b)$.

Definition 38 (General Threshold Function)**.** $Threshold_f : \{-1, 1\}^n \rightarrow \{-1, 1\}$, given a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, is defined as $Threshold_f(x) = sign(f(x))$.

Can we write *OR* as an affine threshold function?

Whether functions are good voting rules:

- **Tribe:**
  - It is a monotone function
  - It is not symmetric
  - It is not odd
  - So in this case the $Aut(f)$ is essentially the blockwise permutations and intrablock permutations. In this case, we can exchange any (i, j) through this permutations. [If i, j are in same block, perform intrablock permutation. If not, then first perform blockwise permutation.]

Let us now move on to the problem of measuring the *significance* of individual bits of the input w.r.t. a boolean function. Formally,

Definition 39 (Sensitive bit)**.** For a function $f : \{-1, 1\}^n \to \{-1, 1\}$, we define $i \in [n]$ to be a sensitive bit for particular $x \in \{-1, 1\}^n$ if $f(x) \neq f(x^{\oplus i})$.

Definition 40 (Influence)**.** For a function $f : \{-1, 1\}^n \to \{-1, 1\}$ and an $i \in [n]$, we define its influence, denoted $Inf_i(f)$ as:

$$Inf_i(f) = \mathbb{P}[f(x) \neq f(x^{\oplus i})]$$

Definition 41 (Edges in Direction of i)**.** Given an hypercube $\{-1, 1\}^n$ and an $i \in [n]$, we define edges in direction of i to be the edges between x and $x^{\oplus i}$.

Definition 42 (Sensitive Edges w.r.t. i for f)**.** Given an hypercube $\{-1, 1\}^n$ and an $i \in [n]$ and a function $f : \{-1, 1\}^n \to \{-1, 1\}$, we define sensitive edges w.r.t. i to be the edges between x and $x^{\oplus i}$ such that $f(x) \neq f(x^{\oplus i}$.

Lemma 43. *For a function $f : \{-1, 1\}^n \to \{-1, 1\}$, we have for any $i \in [n]$,*

$$Inf_i(f) = \frac{\textit{Number of sensitive edges w.r.t. } i \textit{ for } f}{\textit{Number of edges in direction of } i}$$

Proof. Multiply both the numerator and denominator by 2. □

Calculate $Inf_i(Maj_n)$

When does a bit matter? Only when the other bits cause a tie. In that case, we have: $Inf_i(Maj_n) = Inf_1(Maj_n) = \mathbb{P}[x_2 + x_3 + ... + x_n = 0] \sim \sqrt{\frac{2}{\pi n}}$.

Definition 44 (Effect of i in f)**.** Given a function $f : \{-1, 1\}^n \to \{-1, 1\}$ and an $i \in [n]$, we define the effect of i in f, denoted $Eff_i(f)$ to be:

$$Eff_i(f) = \mathbb{P}[f(x) = 1 | x_i = 1] - \mathbb{P}[f(x) = 1 | x_i = -1]$$

Lemma 45. $Eff_i(f) = \hat{f}(\{i\})$

Proof. We know that $\hat{f}(\{i\}) = \left\langle f \, , \, \chi_{\{i\}} \right\rangle$. Also, note that any function f can be written as

$f = 2\mathbf{1}_{f=1} - 1$. Then, we have:

$$\begin{aligned}
\hat{f}(\{i\}) &= \mathbb{E}[f\chi_{\{i\}}] \\
&= \frac{1}{2}\mathbb{E}[f|\chi_{\{i\}} = 1] - \frac{1}{2}\mathbb{E}[f|\chi_{\{i\}} = -1] \\
&= \mathbb{E}[2\mathbf{1}_{f=1}|\chi_{\{i\}} = 1] - \mathbb{E}[2\mathbf{1}_{f=1}|\chi_{\{i\}} = -1] \\
&= Eff_i(f)
\end{aligned}$$

□

Chapter 6

Lecture 6

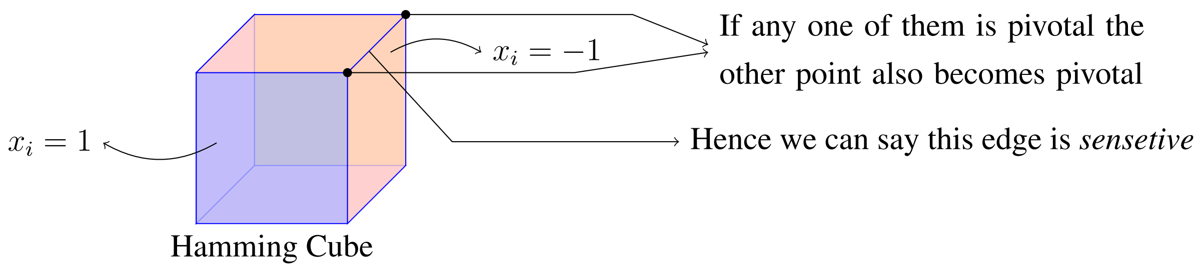
Suppose we have $f : \{\pm 1\}^n \to \{\pm 1\}$.

Definition 46 (Sensitivity)**.** $\forall\, i \in [n]$ and $x \in \{\pm\}^n$ we say i is a sensitive / pivotal coordinate at x for f if

$$f(x) \neq f(x^{\oplus i})$$

where $x^{\oplus i} := (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n)$

Definition 47 (Influence)**.** $\mathrm{Inf}_i[f] = \underset{x \in \{\pm 1\}^n}{Pr}[f(x) \neq f(x^{\oplus i})]$



So $\mathrm{Inf}_i[f] = \frac{\#\text{sensitive edge wrt } i}{\#\text{edges in direction of } i}$.

Lemma 48. $\mathrm{Eff}_i[f] = \hat{f}(\{i\})$

Notation: For $x \in \{\pm 1\}^n$ denote

$$x^{i \to 1} := (x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) \qquad \text{and} \qquad x^{i \to -1} = (x_1, \ldots, x_{i-1}, -1, x_{i+1}, \ldots, x_n)$$

Definition 49 (Partial Derivative in direction of i)**.** For $f : \{\pm 1\}^n \to \{\pm 1\}$ then the partial derivative of f in direction of i is

$$D_i(f)(x) = \frac{f(x^{i\to 1}) - f(x^{i\to -1})}{2}$$

Observation: $D_i f(x) = \begin{cases} \pm 1 & i \text{ is sensitive} \\ 0 & \text{otherwise} \end{cases}$

Lemma 50. *1.* $\mathrm{Inf}_i[f] = \mathbb{E}_x\left[(D_i f)^2\right]$

2. $\mathrm{Eff}_i[f] = \mathbb{E}[D_i f]$

Definition 51 (Total Influence and Effect)**.** For a function $f : \{\pm 1\}^n \to \{\pm 1\}$

- $\mathrm{Inf}[f] := \sum_{i=1}^{n} \mathrm{Inf}_i[f]$
- $\mathrm{Eff}[f] := \sum_{i=1}^{n} \mathrm{Eff}_i[f]$

Fourier Expansion of Derivative

For a function $f : \{\pm 1\}^n \to \{\pm 1\}$

$$D_i f = \sum_{i \notin S \subseteq [n]} \hat{f}(S \cup \{i\}) \chi_S$$

Hence we have $\widehat{D_i f}(S) = \begin{cases} \hat{f}(S \cup \{i\}) & \text{when } i \notin S \\ 0 & \text{otherwise} \end{cases}$

Corollary 52 (Parseval for $D_i f$)**.** $\mathrm{Inf}_i[f] = \sum_{i \in S \subseteq [n]} (\hat{f}(S))^2$

$\mathrm{Inf}_i\,[Parity_n]$

Consider the function $Parity_n : \{\pm 1\}^n \to \{\pm 1\}$ where $Parity_n = x_1 \cdots x_n$ then calculate $\mathrm{Inf}_i[Parity_n] = 1$

Observation: Maximum influence is achieved by $Parity_n$

$\mathrm{Inf}_i\,[Maj_n]$

Consider the function $Maj_n : \{\pm 1\}^n \to \{\pm 1\}$ where $Maj_n = x_1 \cdots x_n$ then calculate $\mathrm{Inf}_i[Maj_n] = \frac{c}{\sqrt{n}}$ for some constant c

$\mathrm{Inf}_i\,[Tribes_{s,w}]$

Consider the function $Tribes_{s,w} = \bigvee_{i=[s]} (\bigwedge_{j\in w} x_{i,j})$ where $s, w \in \mathbb{N}$ calculate $\mathrm{Inf}_i[Tribes_{s,w}] = \frac{1}{2^{w-1}}\left(1 - \frac{1}{2^w}\right)^{s-1}$

Kahn-Kalai-Linial 89

Let $f : \{\pm1\}^n \to \{\pm1\}$ with $\mathrm{Var}[f] = \Omega(1)$ then $\exists\, i \in [n]$ such that $\mathrm{Inf}_i[f] = \Omega\left(\frac{\log n}{n}\right)$

Here

$$\langle f, f\rangle = 1 = \sum_{S\subseteq[n]} (\hat{f}(S))^2 \quad \text{and} \quad \mathrm{Var}[f] = \sum_{S\neq\emptyset, S\subseteq[n]} (\hat{f}(S))^2$$

Definition 53 (Sensitivity of a Point)**.** Let $f : \{\pm1\}^n \to \{\pm1\}$ and $x \in \{\pm1\}^n$. We define $\mathrm{sen}_f(x)$ to be the number of sensitive edges incident to x.

Lemma 54. $\mathrm{Inf}[f] = \mathbb{E}_x[\mathrm{sen}_f(x)]$

Proof.

$$\begin{aligned}
\mathbb{E}_x[\mathrm{sen}_f(x)] &= \frac{1}{2^n}\sum_x \mathrm{sen}_f(x) \\
&= \frac{1}{2^n}\sum_x \sum_{\substack{i\in[n] \\ f(x)\neq f(x^{\oplus i})}} 1 \\
&= \sum_{i\in[n]} \frac{1}{2^n} \sum_{\substack{x \\ f(x)\neq f(x^{\oplus i})}} 1 \\
&= \sum_{i\in[n]} \mathrm{Inf}_i[f] = \mathrm{Inf}[f]
\end{aligned}$$

Another way of seeing this is since each sensitive edge is incident to two points in the boolean cube, $\sum_x \mathrm{sen}_f(x)$ is the twice the #sensitive edges $= 2^n\,\mathrm{Inf}[f]$ □

Lemma 55. $\mathrm{Inf}[f] = \sum_{S\subseteq[n]} |S|(\hat{f}(S))^2$

Proof. From Corollary 52 in $\mathrm{Inf}[f]$ the term $(\hat{f}(S))^2$ appears for all $i \in S$. □

Theorem 56. *Let n is odd. Maj_n is an unique function which maximizes effectiveness.*

Proof. We have $\mathrm{Eff}[f] = \sum_{i=1}^{n} \mathrm{Eff}_i[f]$. Now we also have from Lemma 48 that

$$\mathrm{Eff}_i[f] = \hat{f}(\{i\}) = \langle f, \chi_i \rangle = \langle f, x_i \rangle$$

Therefore we have

$$\mathrm{Eff}[f] = \left\langle f, \sum_{i \in [n]} x_i \right\rangle$$

Therefore the function whose norm is 1 and maximizes $\left\langle f, \sum_{i \in [n]} \right\rangle$ is of same sign with $\sum x_i$. Only f that maximizes will be $sign\left(\sum_{i=1}^{n} x_i\right)$ and notice $sign\left(\sum_{i=1}^{n} x_i\right) = Maj_n$. $\square$

Bibliography

[Arr50] Kenneth Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4), 1950.

[BS23] Thomas F. Bloom and Olof Sisask. The kelley–meka bounds for sets free of three-term arithmetic progressions. *Essential Number Theory*, 2(1):15–44, December 2023.

[KM24] Zander Kelley and Raghu Meka. Strong bounds for 3-progressions, 2024.

[OD14] Ryan O Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[Wol08] Ronald de Wolf. *A Brief Introduction to Fourier Analysis on the Boolean Cube*. Number 1 in Graduate Surveys. Theory of Computing Library, 2008.