High Dimensional Expanders

Analysis of Boolean Functions and New Directions

Instructor: Yuval Filmus *Scribe: Inbal Livni Navon*

1 Introduction to Boolean Functions Analysis

A boolean function is a function $f\colon \{-1,1\}^n \to \{-1,1\}$, and the set $\{-1,1\}^n$ is called the boolean hypercube. In some cases, we consider all functions on the boolean hypercube with range $\mathbb{R}$, i.e. $f\colon \{-1,1\}^n \to \mathbb{R}$.

A fundamental theorem of analysis of boolean functions is the Fourier transform. It is similar to the transform used for continuous functions, only with different basis functions.

Theorem 1.1. *For every $f\colon \{-1,1\}^n \to \mathbb{R}$ there is a unique representation*

$$f(x_1, \dots x_n) = \sum_{S \subset [n]} \hat{f}(S)\chi_S,$$

where $\chi_S = \Pi_{i \in S} x_i$.

The functions χ_S are an orthonormal basis for the collection of functions $\{f\colon \{-1,1\}^n \to \mathbb{R}\}$, with respect to the inner product $\langle f, g\rangle = \mathbb{E}_{x \in \{-1,1\}^n}[f(x)g(x)]$.

We can represent the inner product also by the Fourier coefficients of both functions: $\langle f, g\rangle = \sum_{S \subset [n]} \hat{f}(S)\hat{g}(S)$. This inner product naturally defines the L_2 norm $\|f\|^2 = \langle f, f\rangle = \mathbb{E}_{x \in \{-1,1\}^n}[f^2(x)] = \sum_{S \subset [n]} \hat{f}^2(S)$.

Using the Fourier basis, we can define the degree of a boolean function.

Definition 1.2. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$, the degree of f, denoted by $\deg(f)$, is the degree of the polynomial of its Fourier representation.*

Using this definition, we can split every function $f\colon \{-1,1\}^n \to \mathbb{R}$ to its different Fourier degrees or level, $f = f^{=0} + f^{=1} + \dots + f^{=n}$, where for each $i \in \{0, \dots n\}$, $f^{=i} = \sum_{S \subset [n], |S| = i} \hat{f}(S)\chi_S$.

Definition 1.3. *A function $f\colon \{-1,1\}^n \to \mathbb{R}$ is a d-junta if it depends on at most d variables.*

There is a connection between d-junta to a degree d functions — if f is a d-junta, then its degree is at most d. In addition, every degree d function can be represented as a sum of d-juntas (by taking $\hat{f}(S)\chi_S$ as the d-juntas).

1.1 Linear boolean functions

Suppose f is a boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$, and $\deg(f) \leq 1$. What can we say on f?

The function f has to be either a constant or a dictator (or its negation), because any other linear function has range which is more than $\{-1,1\}$. More explicitly,

$$f \in \{-1, 1, x_i, -x_i\}.$$

Suppose we know that the function $f\colon \{-1,1\}^n \to \{-1,1\}$ is a boolean function, but it is only "almost linear", $\|f^{>1}\|^2 = \|f^{=2} + \dots + f^{=n}\|^2 \leq \epsilon$. Can we say something about such f?

Notice that the function $f^{\leq 1} = f^{=0} + f^{=1}$ is the linear function which is closest (in the L_2 norm defined above) to f. The range of $f^{\leq 1}$ is not $\{-1,1\}$, so it is not boolean.

We can also ask what can be said of a linear function $f'\colon \{-1,1\}^n \to \mathbb{R}$, $\deg(f') = 1$, which is "almost boolean", i.e. $\mathbb{E}_{x\in\{-1,1\}^n}[\mathrm{dist}(f'(x), \{-1,1\})^2] \le \epsilon$.

Can we say that f, f' are close to a dictator?

Theorem 1.4 ([**FKN**])**.** *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be such that $\left\|f^{>1}\right\| \le \epsilon$, then there exists $g\colon \{-1,1\}^n \to \{-1,1\}$, $\deg(g) = 1$ such that $\|f-g\|^2 \le 4\epsilon + O(\epsilon^2)$.*

Let $f'\colon \{-1,1\}^n \to \mathbb{R}$, $\deg(f') = 1$ be such that $\mathbb{E}_{x\in\{-1,1\}^n}[dist(f'(x), \{-1,1\})^2] \le \epsilon$, then there exists $g'\colon \{-1,1\}^n \to \{-1,1\}$, $\deg(g') = 1$ such that $\|f'-g'\|^2 \le 4\epsilon + O(\epsilon^2)$.

Notice that in the case of a boolean function f, $\|f-g\|^2 \le 4\epsilon + O(\epsilon^2)$ implies that $\Pr_{x\in\{-1,1\}^n}[f(x) \ne g(x)] \le \epsilon + O(\epsilon^2)$.

Instead of looking only at functions of degree 1, we can also consider functions of degree at most d. A boolean function of degree at most d can in fact depend on 2^d input variables, as it can be a decision tree of depth d. In a recent paper, [**CHS18**] showed that each degree d boolean function has to be a $C \cdot 2^d$-junta, for a constant $C < 7$. This improves on a classical result of Nisan and Szegedy [**NS94**], who gave an upper bound of $d2^{d-1}$.

1.2 Influence

Definition 1.5. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a boolean function. For every $i \in [n]$, the influence of the i'th variable is*

$$\mathrm{Inf}_i[f] = \Pr_{x\in\{-1,1\}^n}[f(x) \ne f(x^{\oplus i})],$$

where $x^{\oplus i}$ equals x with the ith variable flipped.

The influence can also be written as $\mathrm{Inf}_i[f] = \frac{1}{4}\mathbb{E}_{x\in\{-1,1\}^n}[(f(x) - f(x^{\oplus i}))^2] = \sum_{S\subset[n] \textbf{ s.t } i\in S} \hat{f}^2(S)$.

Definition 1.6. *The average sensitivity or total influence of $f\colon \{-1,1\}^n \to \{-1,1\}$ is the sum of all influences*

$$\mathrm{Inf}[f] = \sum_{i\in[n]} \mathrm{Inf}_i[f].$$

Using the Fourier representation for Inf_i, we can write

$$\mathrm{Inf}[f] = \sum_{i\in[n]} \mathrm{Inf}_i[f] = \sum_{S\subset[n]} |S|\hat{f}^2(S) = \sum_{d\in\{0,\ldots n\}} d\left\|f^{=d}\right\|^2.$$

If we define L to be the following operator: $Lf = \sum_{d\in\{0,\ldots,n\}} df^{=d}$, then $\mathrm{Inf}[f] = \langle Lf, f\rangle$.

The average sensitivity can also be written as

$$\mathrm{Inf}[f] = \mathop{\mathbb{E}}_{x\in\{-1,1\}^n}[\text{number of directions } f(x) \text{ changes value}].$$

Using the above definition, it is easy to see that a d-junta has average sensitivity at most d. There is more than one way to generalize the definition of average sensitivity to non-boolean functions, but the most common one is the one given above.

There is a relation between the average sensitivity and the variance of a boolean function.

Definition 1.7. *Let $f\colon \{-1,1\}^n \to \{-1,1\}$, the variance of f equals $V[f] = \left\|f - \mathbb{E}_{x\in\{-1,1\}^n}[f]\right\|^2$.*

Notice that by definition, $f^{=0} = \mathbb{E}_{x\in\{-1,1\}^n}[f]$, so the variance can be written as $V[f] = \sum_{d\in\{1,\ldots n\}} \left\|f^{=d}\right\|^2$, where we sum over $d \ge 1$. This representation automatically gives us a relation between the variance and the influence:

$$V[f] \le \mathrm{Inf}[f] \le \deg(f)V[f].$$

The left inequality is known as the Poincaré inequality.

Claim 1.8 ([**Friedgut98**]). *A boolean function with low average sensitivity can be approximated by a junta.*

Theorem 1.9 ([**KKL88**]). *Let $f\colon \{-1,1\}^n \to \{-1,1\}$ be a boolean function, then*

$$\max_i \mathrm{Inf}_i[f] = V[f]\Omega\left(\frac{\log n}{n}\right).$$

1.3 Noise Operator

Definition 1.10. *Let $\rho \in [0,1]$ be a constant, then the noise operator T_ρ is defined by*

$$T_\rho f(x) = \mathop{\mathbb{E}}_{y\sim N_\rho(x)}[f(y)],$$

where $y \sim N_\rho(x)$ is distributed as follows, for each $i \in [n]$ independently $y_i = \begin{cases} x_i & w.p.\ \frac{1+\rho}{2} \\ -x_i & w.p.\ \frac{1-\rho}{2} \end{cases}$.

Equivalently, y_i equals x_i with probability ρ, and a random value in $\{-1,1\}$ with probability $1-\rho$.

Using the Fourier basis, we can write $T_\rho f$ by

$$T_\rho f = \sum_{S\subset[n]} \rho^{|S|}\hat{f}(S)\chi_S = \sum_{d\in\{0,\dots n\}} \rho^d f^{=d}.$$

The noise operator can also be described by $T_\rho f = e^{-\log\frac{1}{\rho}L}$, where $Lf = \sum_{d\in\{0,\dots,n\}} df^{=d}$ is as defined previously.

The noise operator "smooths" the function, and lets us bound high norms of f.

Definition 1.11. *For every constant $r > 0$, the r-norm of $f\colon \{-1,1\}^n \to \{-1,1\}$ equals $\|f\|_r = \mathbb{E}[|f|^r]^{\frac{1}{r}}$.*

Claim 1.12. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$, then $\left\|T_{\frac{1}{\sqrt{3}}} f\right\|_4 \le \|f\|_2$, $\left\|T_{\frac{1}{\sqrt{3}}} f\right\|_2 \le \|f\|_{\frac{4}{3}}$ and for any constant $r > 0$, $\left\|T_{\frac{1}{\sqrt{r-1}}} f\right\|_r \le \|f\|_2$, $\left\|T_{\frac{1}{\sqrt{r-1}}} f\right\|_2 \le \|f\|_{r'}$ for r' such that $\frac{1}{r} + \frac{1}{r'} = 1$.*

The proof of the claim is by induction on n. This kind of inequality is known as *hypercontractivity.*

Let $A \subset \{-1,1\}^n$, and let $f = \mathbf{1}_A$, the function which equals 1 on A and 0 else (notice — here the range of f is $\{0,1\}$ and not $\{-1,1\}$). By the claim,

$$\left\|T_{\frac{1}{\sqrt{3}}} f\right\|_2^{\frac{4}{3}} \le \|f\|_{\frac{4}{3}}^{\frac{4}{3}} = \mathop{\mathbb{E}}_{x\in\{-1,1\}^n}[|f(x)|^{\frac{4}{3}}] = \mathop{\mathbb{E}}_{x\in\{-1,1\}^n}[f(x)].$$

Therefore $\left\|T_{\frac{1}{\sqrt{3}}} f\right\|_2^2 \le \mathbb{E}_{x\in\{-1,1\}^n}[f(x)]^{\frac{3}{2}}$.

We can also write the norm of the noise operator by inner product:

$$\left\langle T_{\frac{1}{\sqrt{3}}} f, T_{\frac{1}{\sqrt{3}}} f\right\rangle = \left\langle T_{\frac{1}{3}f}, f\right\rangle = \mathop{\mathbb{E}}_{x\in\{-1,1\}^n}[f(x)T_{\frac{1}{3}}f(x)].$$

Using this notation, we can see that the bound implies that

$$\Pr_{x\in A, y\in N_{\frac{1}{3}}(x)}[y\in A] \le \Pr_{x\in\{-1,1\}}[x\in A]^{\frac{1}{3}}.$$

In other words, if A is small then its expansion in the noisy hypercube is close to 1, a property known as *small set expansion.* Here the noisy hypercube is the weighted graph with vertex set $\{-1,1\}^n$ in which the weight of an edge (x,y) equals $\Pr_{y'\sim N_{\frac{1}{3}}(x)}[y'=y]$.

2 The biased cube

So far, we have only considered the uniform distribution over $\{-1,1\}^n$, and now we are going to see other distributions. The biased cube $(\{0,1\}^n, \mu_p)$ has the vertex set $\{0,1\}^n$ and the p-biased distribution: for each coordinate i independently, $\Pr[x_i = 1] = p$. This changes the definition of inner product accordingly, to $\langle f, g\rangle = \mathbb{E}_{x\sim\mu_p}[f(x)g(x)]$.

In order to use the Fourier representation over the biased cube, we need a different set of basis functions; for every $S \subset [n]$ we use the function $\chi_S = \prod_{i\in S} \frac{x_i - p}{\sqrt{p(1-p)}}$.

Most (though not all) properties of the unbiased cube also hold for the biased cube whenever p is constant. When p is small, the picture changes. The [**FKN**] theorem, for example, no longer holds in the case of small p. We can look at the function $f(x) = x_1 \vee x_2 \vee \cdots \vee x_{\frac{\sqrt{\epsilon}}{p}}$. This function is boolean, and it is ϵ close to being linear. More specifically, it is ϵ-close to the linear function $g(x) = x_1 + x_2 + \cdots + x_{\frac{\sqrt{\epsilon}}{p}}$, which is not boolean. Therefore f is not ϵ-close to a boolean linear function, even though it is boolean and ϵ-close to being linear. It is worth noting that f is $\sqrt{\epsilon}$-close to the constant 0 function.

2.1 Biased cube and the $G(n,p)$ model

The biased cube can represent the edge set in the $G(n,p)$ model. We set $n' = \binom{n}{2}$, and a function $f\colon \{0,1\}^{n'} \to \{0,1\}$ represents a graph property for a graph with n vertices.

For example, suppose $f(G) = 1$ if and only if the graph G contains a triangle. The expected number of triangles in a graph in the $G(n,p)$ model is $p^3\binom{n}{3}$, so if we set $p = \frac{c}{n}$ the expected number of triangles is $\frac{c^3}{6}$, and

$$\Pr_{G\sim G(n,\frac{c}{n})}[f(G) = 1] \approx e^{-\frac{c^3}{6}}.$$

The critical probability for the existence of a triangle is $p = \frac{c}{n}$, and the relevant scale is also $\frac{1}{n}$ (increasing p by a factor of $\frac{1}{n}$ changes the probability by a constant).

In contrast, suppose $f(G) = 1$ if G is connected, in this case it is known that

$$\Pr_{G\sim G(n,\frac{\log n + c}{n})}[f(G) = 1] \approx e^{-e^{-c}}.$$

In this case, the critical probability is $\frac{\log n}{n}$, but notice that the scale is still $\frac{1}{n}$, because if we increase p by $\frac{c}{n}$ to p then the probability that the graph is connected changes by a constant. This is called a sharp threshold, in which the scale is much smaller than the critical probability.

The Russo–Margulis lemma shows that for monotone properties f,

$$\frac{\partial \Pr_{G\sim G(n,p)}[f(G) = 1]}{\partial f} = \mathrm{Inf}[f]_p.$$

Therefore, if the threshold is coarse (not sharp), then there is a point near the critical probability at which the total influence is small. When p is constant, Friedgut's junta theorem [**Friedgut98**] shows that f is close to a junta. When p is small, sharp threshold theorems such as Friedgut's [**Friedgut99**] reveal a more complicated "local junta" structure, as demonstrated by the property of containing a triangle (here the "local junta" structure is related to the fact that the number of triangles is expected to be small when p is close to the critical probability).

3 The Slice

In both the uniform distribution and the biased distribution, the coordinates were independent, but it does not have to be the case. For example, we can look at the slice, which is also called the Johnson scheme.

$$\binom{[n]}{k} = \left\{ (x_1, \dots, x_n) \in \{0,1\}^n \;\middle|\; \sum_{i\in[n]} x_i = k \right\} = J(n,k).$$

In many aspects, this scheme is similar to the biased cube with $k = n \cdot p$, but not always.

In the slice there can be more than a single polynomial representation to the constant 0 function. Both $f(x) = 0$ and $f(x) = \sum_{i\in[n]} x_i - k$ are polynomials that represents the constant 0 function on the slice. This implies that there is more than a single polynomial representation for every function $f\colon \binom{[n]}{k} \to \mathbb{R}$. The following theorem shows that there is a unique representation if we restrict the class of polynomials.

Theorem 3.1 (Dunkl)**.** *Every* $f\colon \binom{[n]}{k} \to \mathbb{R}$ *has a unique representation as a multilinear polynomial* P *of degree at most* $\min\{k, n-k\}$ *which satisfies* $\sum_{i\in[n]} \frac{\partial P}{\partial x_i} = 0$.

A polynomial P satisfying $\sum_{i\in[n]} \frac{\partial P}{\partial x_i} = 0$ is called *harmonic*. Examples of harmonic multilinear polynomials are $1, x_1 - x_2, (x_1 - x_2)(x_3 - x_4)$. These examples are representative in the sense that *every* harmonic multilinear polynomial is a linear combination of such polynomials.

This theorem gives a partition of each $f\colon \binom{[n]}{k} \to \mathbb{R}$ into levels as before $f = f^{=0} + \dots + f^{=k}$, where $f^{=d}$ is a polynomial of degree d. The different levels of f are orthogonal to each other. We can also get the operators L, T_ρ:

$$Lf = \sum_{d\in\{0,\dots,k\}} d\left(1 - \frac{d-1}{n}\right) f^{=d}; \quad T_\rho f = \sum_{d\in\{0,\dots,k\}} \rho^{d(1-\frac{d-1}{n})} f^{=d},$$

in the definition we assume that $k \le n - k$, else $d \in \{0, \dots n-k\}$.

The theorem does not gives us a Fourier basis. In order to get a Fourier basis, we need an orthogonal basis for the set of polynomials P such that $\sum_{i\in[n]} \frac{\partial P}{\partial x_i} = 0$. This basis has to be smaller than the basis in the uniform case. The Fourier basis χ_S in the uniform case has $\binom{n}{d}$ degree d polynomials, and in this case only $\binom{n}{d} - \binom{n}{d-1}$ are required.

There is a Fourier basis for the Johnson scheme $J(n,k)$, known as the Gelfand–Tsetlin (GZ) basis, which consists of common eigenvectors of the Young–Jucys–Murphy elements $(1\ m) + \dots + (m-1\ m)$ for all m. These are elements of the group algebra of S_n, and they act on polynomials by transposing pairs of variables and summing the results. For example,

$$(x_1 - x_2)^{(1\,3)+(2\,3)} = (x_1 - x_2)^{(1\,3)} + (x_1 - x_2)^{(2\,3)} = (x_3 - x_2) + (x_1 - x_3) = x_1 - x_2.$$

For $d = 1$ the basis polynomials includes: $x_2 - x_1, 2x_3 - x_1 - x_2, 3x_4 - x_1 - x_2 - x_3$ (these are not normalized to have unit norm). When $d > 1$ the polynomials are a bit more complicated, but there is an explicit formula for them due to Filmus.

In many cases, it is enough to use the partition into levels and the uniqueness of the harmonic polynomials, without needing to use the Fourier basis itself.

In the case of the slice, we can't define Inf_i as before, because if $x \in \binom{[n]}{k}$, then $x^{\oplus i} \notin \binom{[n]}{k}$. Instead, we define $\mathrm{Inf}_{i,j}$, which is influence in the direction i, j in which we switch x_i with x_j. We can then define Inf_i as the sum of $\mathrm{Inf}_{i,j}$ for all $j \ne i$. Equivalently (using a different normalization), we can define Inf_i using the operation which switches x_i with a random coordinate.

4 Grassmann Scheme

In the Grassmann scheme, our domain is $\binom{[n]}{k}_q = J_q(n,k) = Gr_q(n,k)$, which consists of all k-dimensional subspaces in $GF(q)^n$, for a prime power q. There is a generalization of Dunkl's theorem for the Grassmann scheme.

Every k-dimensional subspace $V \in Gr_q(n,k)$ has a dual $(n-k)$-dimensional subspace $V^\perp \in Gr_q(n,n-k)$, and this gives an isomorphism $Gr_q(n,k) \equiv Gr_q(n,n-k)$ which is analogous to the trivial isomorphism $J(n,k) \equiv J(n,n-k)$; however, the operation $V \mapsto V^\perp$ is much less trivial than the corresponding operation $S \mapsto \overline{S}$ on the slice.

Let us fix $q = 2$. How does a boolean function of degree 1 look like in the Grassman scheme? A function $f\colon Gr_2(n,k) \to \mathbb{R}$ has degree 1 if $f(V) = \sum_{x \in GF(2)^n \setminus \{0\}} \alpha_x [x \in V]$, when $[x \in V]$ equals 1 if $x \in V$ and 0 else. Equivalently (though this is not obvious), it has degree 1 if $f(V) = \sum_{y \in GF(2)^n \setminus \{0\}} \beta_y [y \perp V]$.

When can such a function be boolean? If almost all of the α_x or the β_y equal 0, i.e., if $f \in \{0, 1, [x \in V], [x \notin V], [y \perp V], [y \not\perp V], [x \in V \text{ or } y \perp V], [x \notin V \text{ and } y \not\perp V]\}$, for x, y such that $x \not\perp y$. This theorem has a long inductive proof, and no simple proof is known at present (though probably one exists).

5 Other Domains

In this section we briefly describe a few other domains.

- Another domain, which in some sense encompasses the Grassmann scheme, is the bilinear scheme, in which the inputs are matrices over $GF(q)$.
- It is also possible to consider a layer in a high dimensional expander as a domain. In this case our vertex set is a small subset of the Johnson scheme with good expansion properties.
- It is also possible to preform boolean function analysis on the symmetric group S_n.