1 Boolean functions and Walsh-Fourier system

In this chapter we would like to study boolean functions, namely functions $f : \{-1,1\}^n \to \{-1,1\}$, using methods of harmonic analysis. Recall that the discrete cube $\{-1,1\}^n$ is equipped with several structures. One of them is a graph structure. The points $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are neighbours if and only if $|\{1 \le i \le n : \ x_i \ne y_i\}| = 1$. It means that x and y differ only on one coordinate. It this case if $y = (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_n)$, so the difference is on i-th coordinate, we write $y = x^i$. We also write

$$f_i(x) = f(x) - f(x^i).$$

Another important structure is a structure of measure space. Of course we can equipped $\{-1,1\}^n$ with many different measure, but the most important one is the uniform measure,

$$\mu(S) = \frac{1}{2^n}|S|, \quad S \subset \{-1,1\}^n.$$

Having a measure μ on a discrete cube and a function $f : \{-1,1\}^n \to \mathbb{R}$ we can consider the expectation of f,

$$\mathbb{E}f = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)$$

and the L_p norm

$$\|f\|_p = \left(\mathbb{E}|f|^p\right)^{1/p}, \quad p > 0.$$

We write $\mathbb{P}(A) = \mathbb{E}\mathrm{I}_A$. We also have a structure of a Hilbert space $L_2(\{-1,1\}^n, \mu)$ of all functions $f : \{-1,1\}^n \to \mathbb{R}$ with a scalar product

$$\langle f, g\rangle = \mathbb{E}fg = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)g(x).$$

The space $L_2(\{-1,1\}^n, \mu)$ has dimension 2^n and the functions

$$\delta_y(x) = \begin{cases} 1 & x = y \\ 0 & x \ne y \end{cases}$$

form the basis of this space. It is an orthogonal basis. However, we have another basis, which we will frequently use. Let $[n] = \{1, \ldots, n\}$. Namely, we define

$$w_S(x_1, \ldots, x_n) = \prod_{i \in S} x_i, \quad S \subset [n], \quad w_\emptyset \equiv 1.$$

We have $w_S \cdot w_T = w_{S\Delta T}$. The measure μ is a product measure, therefore

$$\mathbb{E}x_{i_1} \cdot \ldots \cdot x_{i_k} = \mathbb{E}x_{i_1} \cdot \ldots \cdot \mathbb{E}x_{i_k} = 0.$$

It follows that

$$\mathbb{E}w_S = \begin{cases} 1 & S = \emptyset \\ 0 & S \neq \emptyset \end{cases}, \qquad \mathbb{E}w_{S\Delta T} = \begin{cases} 1 & S = T \\ 0 & S \neq T \end{cases}.$$

Therefore $(w_S)_{s\subset[n]}$ is an orthonormal basis and every function can be written in the form

$$f = \sum_{S\subset[n]} a_S w_S,$$

where $(a_s)_{s\subset[n]}$ are some real coefficients. We have

$$\langle f, w_T\rangle = \left\langle \sum_{S\subset[n]} a_s w_S, w_T \right\rangle = \sum_{S\subset[n]} a_S \langle w_S, w_T\rangle = a_T,$$

thus

$$f = \sum_{S\subset[n]} \langle f, w_S\rangle w_S.$$

Sometimes we write $a_S = \hat{f}(S)$.

The discrete cube possess a graph structure, namely for $x, y \in \{-1,1\}^n$ the point x is a neighbour of y (which will be denoted by $x \sim y$) if and only if there exists $1 \leq i \leq n$ such that $y = x^i$.

2 Influences of boolean function

Let $v \in \{-1,1\}^n$ and let $f : \{-1,1\}^n \to \{-1,1\}$. We define the *sensitivity* of v by

$$s(v,f) = |\{1 \leq i \leq n : \ f(v^i) \neq f(v)\}|.$$

The *average sensitivity* is simply

$$as(f) = \mathbb{E}s(f) = \int s(v,f) \, \mathrm{d}\mu(y).$$

The *influence* of the i-th variable is defined as

$$I_i(f) = \mathbb{P}(f(x) \neq f(x^i)) = \frac{1}{2^n} \left|\left\{x \in \{-1,1\}^n : \ f(x) \neq f(x^i)\right\}\right|.$$

In other word, I_i is the probability that the value of f is undefined if we assigned values to x_j for $i \neq j$. The randomness is with respect to the assignment of the values of x_j.

We prove that the sum of the influences is equal to the average sensitivity. Indeed, we have

$$\sum_{i=1}^n I_i(f) = \frac{1}{2^n} \sum_{i=1}^n |\{x : \ f(x) \neq f(x^i)\}| = \sum_{i=1}^n \int \mathrm{I}_{\{x:\, f(x) \neq f(x^i)\}}(y) \, \mathrm{d}\mu(y)$$
$$= \int \sum_{i=1}^n \mathrm{I}_{\{x:\, f(x) \neq f(x^i)\}}(y) \, \mathrm{d}\mu(y) = \int s(y,f) \, \mathrm{d}\mu(y) = as(f).$$

There is an one-to-one correspondence between boolean functions and subsets of the discrete cube. Namely, if $f : \{-1,1\}^n \to \{-1,1\}$ then we can define $A_f = \{x : f(x) = 1\}$. If $A \subset \{-1,1\}^n$ then we also have $f_A(x) = 2\mathrm{I}_A(x) - 1$. If we have sets $A, B \subset \{-1,1\}^n$ with then we define

$$E(A,B) = |\{(a,b) : \ a \in A, b \in B, a \sim b\}|.$$

The quantity $E(A, A^c)$ is the so-called *edge boundary* of A. We have

$$\frac{|E(A,A^c)|}{2^{n-1}} = \frac{2|E(A,A^c)|}{2^n} = \frac{\sum_{i=1}^n |\{x : \ f_A(x) \neq f_A(x^i)\}|}{2^n} = \sum_{i=1}^n I_i.$$

We are now ready to give a crucial definition in this chapter.

Definition 1. The influence (total influence) of a boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is defined as

$$I(f) = \sum_{i=1}^n I_i = \mathbb{E}s(f) = \frac{|E(A,A^c)|}{2^{n-1}}.$$

3 Examples of boolean functions and their influences

In this section we analyse some basis examples of boolean functions.

- Dictator: $\mathrm{Dict}_n(x_1, \ldots, x_n) = x_j$, $1 \leq j \leq n$,

  Clearly, we have

$$I_i(\mathrm{Dict}_n) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}, \qquad I(\mathrm{Dict}_n) = 1, \qquad \mathbb{E}(\mathrm{Dict}_n) = 0.$$

- Junta (k-junta): $f(x_1, \ldots, x_n) = g(x_{i_1}, \ldots, x_{i_k})$, where $g : \{-1,1\}^k \to \{-1,1\}$ and $1 \le k < n$.

- Parity: $\mathrm{Par}_n(x_1, \ldots, x_n) = x_1 \cdot \ldots \cdot x_n$. Note that Parity is equal to the Walsh function of highest degree, namely $w_{[n]}$.

$$I_i(\mathrm{Par}_n) = 1, \qquad I(\mathrm{Par}_n) = n, \qquad \mathbb{E}(\mathrm{Par}_n) = 0.$$

- Majority: $\mathrm{Maj}_n(x_1, \ldots, x_n) = \mathrm{sgn}(x_1 + \ldots + x_n)$, n is odd,

$$I_i(\mathrm{Maj}_n) = \frac{1}{2^{n-1}} \binom{n-1}{\frac{n-1}{2}} = O\left(\frac{1}{\sqrt{n}}\right), \quad I(\mathrm{Maj}_n) = \frac{n}{2^{n-1}} \binom{n-1}{\frac{n-1}{2}} = O(\sqrt{n}),$$
$$\mathbb{E}(\mathrm{Maj}_n) = 0.$$

- AND: $\mathrm{AND}_n(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n)$,

$$I_i(\mathrm{AND}_n) = \frac{1}{2^{n-1}}, \qquad I(\mathrm{AND}_n) = \frac{n}{2^{n-1}}, \qquad \mathbb{E}(\mathrm{AND}_n) = -1 + \frac{1}{2^{n-1}}.$$

- OR: $\mathrm{OR}_n(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n)$

$$I_i(\mathrm{OR}_n) = \frac{1}{2^{n-1}}, \qquad I(\mathrm{OR}_n) = \frac{n}{2^{n-1}}, \qquad \mathbb{E}(\mathrm{OR}_n) = 1 - \frac{1}{2^{n-1}}.$$

- Tribes: take $n = mk$ and divide n variables into m groups (tribes), each of cardinality k. The value of our function is 1 if and only if there exists a tribe which says 'yes'. The tribe says 'yes' if all values of spines in this tribe is 1. So the Tribes function is OR of ANDs. We can write

$$\mathrm{Tribes}_{k,m}(x_1, \ldots, x_n) = OR\left(AND(x_1, ..., x_k), ..., AND(x_{(m-1)k+1}, ..., x_{mk})\right).$$

  To calculate I_i observe that if x_i wants to decide then others variables in its tribe has to take value 1 and in $m-1$ other tribes there must be at least 1 variable with value 0 in each tribe. Therefore,

$$I_i(\mathrm{Tribes}_{k,m}) = \frac{1}{2^{k-1}} \left(1 - \frac{1}{2^k}\right)^{m-1}, \quad I(\mathrm{Tribes}_{k,m}) = \frac{km}{2^{k-1}} \left(1 - \frac{1}{2^k}\right)^{m-1},$$
$$\mathbb{E}(\mathrm{Tribes}_{k,m}) = 1 - 2\left(1 - \frac{1}{2^k}\right)^m.$$

Now we would like to find the value $k = k(n)$ for which $\mathbb{P}(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}) = p$. Let us take

$$k(n) = \log_2\left(\frac{n}{-\ln(1-p)}\right) - \log_2\log_2 n.$$

Of course $k(n)$ and $n/k(n)$ should be integers, but who cares... Since for a boolean function f we have $\mathbb{E}f = 2\mathbb{P}(f=1) - 1$, therefore

$$\begin{aligned} 1 - \mathbb{P}(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}} = 1) &= \left(1 - \frac{1}{2^{k(n)}}\right)^{n/k(n)} \\ &= \left(1 + \frac{(\ln(1-p))(\log_2 n)}{n}\right)^{n/k(n)}. \end{aligned}$$

Let

$$a_n = \frac{n}{(\ln(1-p))(\log_2 n)}.$$

Clearly, $\lim_{n\to\infty}|a_n| = +\infty$. Therefore $\lim_{n\to\infty}(1+\frac{1}{a_n})^{a_n} = e$. Moreover,

$$\lim_{n\to\infty}\frac{n}{k(n)a_n} = \lim_{n\to\infty}\frac{(\ln(1-p))(\log_2 n)}{\log_2\left(\frac{n}{-\ln(1-p)}\right) - \log_2\log_2 n} = \ln(1-p).$$

It follows that

$$\lim_{n\to\infty}\mathbb{P}(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}} = 1) = 1 - e^{\ln(1-p)} = p.$$

Let us now calculate the asymptotic behaviour of $I_i(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}})$. We have

$$\begin{aligned} I_i(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}) &= \frac{1}{2^{k(n)-1}}\left(1 - \frac{1}{2^k}\right)^{n/k(n)-1} \\ &= \frac{1}{2^{k(n)-1}}\left(1 - \frac{1}{2^k}\right)^{-1}\Big(1 - \mathbb{P}(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}} = 1)\Big) \\ &\approx \frac{1}{2^{k(n)-1}}(1-p) \approx 2(1-p)\ln\left(\frac{1}{1-p}\right)\frac{\log_2 n}{n}. \end{aligned}$$

Therefore,

$$I_i(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}) \approx 2(1-p)\ln\left(\frac{1}{1-p}\right)\frac{\log_2 n}{n}, \quad n\to\infty,$$

$$I_i(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}) \approx 2(1-p)\ln\left(\frac{1}{1-p}\right)\log_2 n, \quad n\to\infty.$$

If $p \le 1/2$ then we have

$$I_i(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}) \le Cp\frac{\log_2 n}{n}.$$

4 Basic estimates of $I(f)$

We would like to make a connection between classical isoperimetric inequalities an inequalities in for the discrete cube. We are going to prove the following proposition

Proposition 1. Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $\mu(f) = \mathbb{P}(f = 1)$. Then for $\mu(f) \le 1/2$ we have

$$I(f) \ge 2\mu(f)\ln\left(\frac{1}{\mu(f)}\right).$$

We first prove the following lemma.

Lemma 1 (Loomis-Whitney inequality)**.** Let $A \subset \mathbb{R}^n$ be an open set in $\mathbb{R}^n$ and let $P_i : \mathbb{R}^n \to \mathbb{R}^{n-1}$ be a projection given by $P_i(x_1,\dots,x_n) = (x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)$. Then

$$|A|^{n-1} \le |P_1(A)|\cdot\ldots\cdot|P_n(A)|.$$

To prove this we need an elementary inequality.

Lemma 2 ($G(A) \ge A(G)$ inequality)**.** Consider an array of nonnegative numbers $(a_{i,j})_{i,j=1}^{n,m}$. Then compute the geometric mean of each row and the arithmetic mean of each column. Therefore, we have a diagram

$$\begin{array}{ccccccc}
 & a_{11} & a_{12} & \dots & a_{1n} & \to & G_1 = \sqrt[n]{a_{11}\cdot\ldots a_{1n}} \\
 & a_{21} & a_{22} & \dots & a_{2n} & \to & G_2 = \sqrt[n]{a_{11}\cdot\ldots a_{1n}} \\
 & \vdots & \vdots & \ddots & \vdots & & \\
 & a_{m1} & a_{22} & \dots & a_{mn} & \to & G_m = \sqrt[n]{a_{11}\cdot\ldots a_{1n}} \\
 & \downarrow & \downarrow & & \downarrow & & \\
 & A_1 = \frac{a_{11}+\ldots+a_{m1}}{m} & A_2 = \frac{a_{12}+\ldots+a_{m2}}{m} & \dots & A_n = \frac{a_{1n}+\ldots+a_{mn}}{m} & &
\end{array}$$

Then the geometric mean of the arithmetic means of columns is not less then the arithmetic mean of the geometric means of rows, namely

$$\sqrt[n]{A_1A_2\cdot\ldots\cdot A_n} \ge \frac{G_1+G_2+\ldots+G_m}{m}.$$

It other words

$$\prod_{i=1}^{n}\left(\sum_{j=1}^{m} a_{ij}\right)^{1/n} \geq \sum_{j=1}^{m}\left(\prod_{i=1}^{n} a_{ij}\right)^{1/n}.$$

Proof. Using A-G inequality we obtain

$$\sum_{i=1}^{n}\frac{a_{ji}}{A_i} \geq n\cdot\sqrt[n]{\prod_{i=1}^{n}\frac{a_{ji}}{A_i}} = \frac{nG_j}{\sqrt[n]{A_1A_2\ldots A_n}}, \quad 1\leq i\leq m.$$

Adding this inequalities we obtain

$$\sum_{j=1}^{m}\sum_{i=1}^{n}\frac{a_{ji}}{A_i} \geq \sum_{j=1}^{m}\frac{nG_j}{\sqrt[n]{A_1A_2\ldots A_n}} = nm\frac{\frac{G_1+G_2+\ldots+G_m}{m}}{\sqrt[n]{A_1A_2\ldots A_n}}.$$

Since

$$\sum_{j=1}^{m}\sum_{i=1}^{n}\frac{a_{ji}}{A_i} = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{a_{ji}}{A_i} = \sum_{i=1}^{n}\frac{mA_i}{A_i} = mn,$$

we obtain

$$\sqrt[n]{A_1A_2\cdot\ldots\cdot A_n} \geq \frac{G_1+G_2+\ldots+G_m}{m}.$$

□

Proof of Lemma 1. It suffices to prove the following discrete version of this theorem. Namely, consider a partition of $\mathbb{R}^n$ into cubes of size $\delta\times\ldots\times\delta$,

$$\mathbb{R}^n = \bigcup_{k_1,\ldots,k_n\in\mathbb{Z}}[\delta k_1, \delta(k_1+1)]\times\ldots\times[\delta k_n, \delta(k_n+1)]$$

This will be called a δ-partition. Consider a set of N cubes, where each cube is an element of this partition. If project our cubes using P_i, we obtain a new set cubes in the partition of $\mathbb{R}^n$. Some of the cubes may be projected onto the same cube. Let N_i be the number of cubes after projecting. Then

$$N^{n-1} \leq N_1N_2\cdot\ldots\cdot N_n.$$

Having this discrete version we now prove that this implies the Loomis-Whitney inequality. For every $\varepsilon>0$ there exists $\delta>0$ such that there exists a set $\tilde{A}\subset A$ which is a sum of N cubes in the δ-partition of $\mathbb{R}^n$, such that $|A\backslash\tilde{A}|<\varepsilon$. We have

$$|\tilde{A}|^{n-1} = N^{n-1}\delta^{n(n-1)} \leq (N_1\delta^{n-1})\cdot\ldots\cdot(N_n\delta^{n-1}) \leq |P_1(A)|\cdot\ldots\cdot|P_n(A)|.$$

Now it suffices to take $\varepsilon \to 0$ and observe that $|\tilde{A}| \to |A|$.

Now we prove our discrete version. We use induction. For $n = 2$ the assertion is trivial. Let us project our cubes onto the first coordinate. We obtain elements $I_1, \ldots, I_k$ of the δ-partition of $\mathbb{R}$. Let $T_1, T_2, \ldots, T_k$ be the sets of cubes that are projected onto $I_1, I_2, \ldots, I_k$, respectively. One can project the cubes from T_i onto $\mathbb{R}^{n-1}$ using P_j and obtain the sets T_{ij} of cubes in δ-partition of $\mathbb{R}^{n-1}$. Let a_i be the cardinality of T_i and let a_{ij} be the cardinality of T_{ij}. We have some rather trivial relations,

$$\sum_{i=1}^{k} a_i = N, \quad \sum_{i=1}^{k} a_{ij} = N_j, \quad a_i \le N_1.$$

The inequality $a_i \le N_1$ follows from the fact that two different cubes with the same projection onto the linear subspace $V = \mathrm{Lin}(e_1)$ must have different projection onto the complement of V (the cube is a product of these two projections). From the induction hypothesis we have

$$a_i^{n-2} \le a_{i2} \cdot \ldots \cdot a_{in}, \quad i = 1, \ldots, k.$$

Combining this with $a_i \le N_1$ we obtain $a_i^{n-1} \le N_1 \cdot a_{i2} \cdot \ldots \cdot a_{in}$. Therefore, using $G(A) \ge A(G)$ inequality

$$\begin{aligned} N = \sum_{i=1}^{n} a_i \le \sum_{i=1}^{n} (N_1 \cdot a_{i2} \cdot \ldots \cdot a_{in})^{1/(n-1)} &= N_1^{1/(n-1)} \sum_{i=1}^{k} \left(\prod_{j=2}^{m} a_{ij} \right)^{1/(n-1)} \\ \le N_1^{1/(n-1)} \prod_{j=2}^{m} \left(\sum_{i=1}^{k} a_{ij} \right)^{1/(n-1)} &= \prod_{j=1}^{m} N_j^{1/(n-1)} \end{aligned}$$

This finishes the proof. □

Now we are ready to prove Proposition 1.

Proof. Consider the following family $\mathcal{C}$ of cubes in $[0,1]^n$,

$$\mathcal{C}_{\varepsilon_1,\ldots,\varepsilon_n} = \left[\frac{\varepsilon_1}{2}, \frac{1}{2} + \frac{\varepsilon_1}{2} \right] \times \ldots \times \left[\frac{\varepsilon_n}{2}, \frac{1}{2} + \frac{\varepsilon_n}{2} \right], \qquad \varepsilon_1, \ldots, \varepsilon_n \in \{0,1\}.$$

Now we define a subset $A = A_f \subset [0,1]^n$ which is an union of some cubes from $\mathcal{C}$ by the following rule: $\mathcal{C}_{\varepsilon_1,\ldots,\varepsilon_n} \subset A$ if and only if $f(2\varepsilon_1 - 1, \ldots, 2\varepsilon_n - 1) = 1$. Clearly $\mu(f) = |A|$. Let us fix $1 \le i \le n$. We have 2^{n-1} pairs

$$\left(\mathcal{C}_{\varepsilon_1,\ldots,\varepsilon_{i-1},0,\varepsilon_{i+1},\ldots,\varepsilon_n}, \mathcal{C}_{\varepsilon_1,\ldots,\varepsilon_{i-1},1,\varepsilon_{i+1},\ldots,\varepsilon_n} \right), \qquad \varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_{i+1}, \ldots, \varepsilon_n \in \{0,1\}.$$

Suppose a is a number of pairs such that both cubes are not contained in A, b is a number of pair such that both cubes contained in A and let c be a number of pairs such that one of the cubes is contained in A and another one is not. We have

$$\mu(f) = \mu(f) = \frac{b}{2^{n-1}} + \frac{c}{2^n}, \quad I_i = I_i(f) = \frac{c}{2^{n-1}}, \quad |P_i(A)| = \frac{b+c}{2^{n-1}}.$$

Therefore

$$|P_i(A)| = \mu(f) - \frac{I_i}{2} + I_i = \mu(f) + \frac{I_i}{2}, \qquad i = 1, \ldots, n.$$

□

From the Lemma 1 we have

$$\mu(f)^{n-1} = |A|^{n-1} \leq |P_1(A)| \cdot \ldots \cdot |P_n(A)| = \left(\mu(f) + \frac{I_1}{2}\right) \cdots \left(\mu(f) + \frac{I_n}{2}\right),$$

thus

$$\frac{1}{\mu(f)} \leq \left(1 + \frac{I_1}{2\mu(f)}\right) \cdots \left(1 + \frac{I_n}{2\mu(f)}\right)$$

and therefore

$$\ln\left(\frac{1}{\mu(f)}\right) \leq \ln\left(1 + \frac{I_1}{2\mu(f)}\right) + \ldots + \ln\left(1 + \frac{I_n}{2\mu(f)}\right) \leq \frac{I_1 + \ldots + I_n}{2\mu(f)} = \frac{I(f)}{2\mu(f)}.$$

It follows that

$$I(f) \geq 2\mu(f) \ln\left(\frac{1}{\mu(f)}\right).$$

We would like to prove a better bound. Namely, in the above estimate one can take $\log_2$ instead of ln.

Proposition 2. Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $\mu(f) = \mathbb{P}(f = 1)$. Then for $\mu(f) \leq 1/2$ we have

$$I(f) \geq 2\mu(f) \log_2\left(\frac{1}{\mu(f)}\right).$$

Hence, if $\mu(f) = 1/2$ then we have $I(f) \geq 1$. This last inequality is optimal since $I(\mathrm{Dict}_n) = 1$ and $\mu(\mathrm{Dict}_n) = 1/2$.

It suffices to prove the following lemma.

Lemma 3. Let $A \subset \{-1,1\}^n$, $|A| = m$. Then $|E(A, A^c)| \geq m(n - \log_2 m)$.

Indeed, this lemma implies Proposition 2. Take $A = A_f$ and note that $\mu(f) = \frac{|A|}{2^n} = \frac{m}{2^n}$. Therefore

$$I(f) = \frac{|E(A, A^c)|}{2^{n-1}} \geq \frac{m(n - \log_2 m)}{2^{n-1}} = \frac{m}{2^{n-1}}(n - \log_2(2^n \mu(f))) = -2\mu(f)\log_2 \mu(f).$$

To prove Lemma 3 we prove

Lemma 4. Let $A \subset \{-1, 1\}^n$. Let $v \in A$. Take $d_A(v) = |\{u \in A : \ u \sim v\}|$. Then

$$|A| \geq 2^{\bar{d}}, \qquad \text{where} \quad \bar{d} = \frac{\sum_{v \in A} d_A(v)}{|A|}.$$

This lemma implies Lemma 3. Indeed,

$$\begin{aligned} |E(A, A^c)| &= |\{(v, u) : \ v \in A, u \in A^c, v \sim u\}| = \sum_{v \in A} |\{u : \ u \in A^c, u \sim v\}| \\ &= \sum_{v \in A} (n - |\{u : \ u \in A, u \sim v\}|) = \sum_{v \in A} (n - d_A(v)) = n|A| - \bar{d}|A|. \end{aligned}$$

If $m = |A|$ then $m \geq 2^{\bar{d}}$. Thus $\bar{d} \leq \log_2 m$. We arrive at

$$|E(A, A^c)| = |A|(n - \bar{d}) = m(n - \bar{d}) \geq m(n - \log_2 m).$$

We are to prove Lemma 4.

Proof o Lemma 4. It is easy to check that for $n = 1$ our assertion is true. We use induction. Divide $\{-1, 1\}^n$ into two subcubes of dimension $n - 1$, $\{x_1 = -1\}$ and $\{x_1 = 1\}$. Consider

$$A_1 = A \cap \{x_1 = -1\}, \qquad A_2 = A \cap \{x_1 = 1\}.$$

Let $m_1 = |A_1|$ and $m_2 = |A_2|$. Without loss of generality we can assume that $0 \leq m_1 \leq m_2$. Let s be the number of vertices between A_1 and A_2. Clearly $s \leq m_1$. For $i = 1, 2$, using Lemma 3 we have

$$m_i \log_2 m_i \geq \sum_{v \in A_i} d_{A_i}(v) = \left(\sum_{v \in A_i} d_G(v) \right) - s.$$

We use the notation $0 \log_2 0 = 0$. Summing this inequalities we obtain

$$m_1 \log_2 m_1 + m_2 \log_2 m_2 \geq \left(\sum_{v \in A} d_A(v) \right) - 2s \geq \left(\sum_{v \in A} d_A(v) \right) - 2m_1.$$

Our goal in to prove

$$(m_1 + m_2) \log_2(m_1 + m_2) \geq \sum_{v \in A} d_A(v).$$

If suffices to check that

$$(m_1 + m_2) \log_2(m_1 + m_2) \geq m_1 \log_2 m_1 + m_2 \log_2 m_2 + 2m_1, \qquad 0 \leq m_1 \leq m_2.$$

We state this inequality as lemma.

Lemma 5. Let $0 \leq x \leq y$. Then

$$(x + y) \log_2(x + y) \geq x \log_2 x + y \log_2 y + 2x.$$

Proof. The inequality is true for $x = 0$. Therefore we can assume $x > 0$. Take $\gamma = y/x$. We have

$$\begin{aligned}
(x + y) \log_2(x + y) - x \log_2 x - y \log_2 y &= x \log_2 x(1 + \gamma) + y \log_2 y(1 + 1/\gamma) \\
&= x \log_2(1 + \gamma) + y \log_2(1 + 1/\gamma) = x \log_2(1 + \gamma) + x\gamma \log_2(1 + 1/\gamma) \\
&\geq x \log_2(1 + \gamma) + x \log_2(1 + 1/\gamma) = x \log_2((1 + \gamma)(1 + 1/\gamma)) \\
&= x \log_2(\gamma + 1/\gamma + 2) \geq x \log_2(2 + 2) = 2x.
\end{aligned}$$

□

Lemma 4 follows. □

5 Parseval's identity

Recall that we can always write

$$f = \sum_{s \subset [n]} a_s w_S,$$

where $(w_S)_{s \subset [n]}$ are the so-called Walsh functions. Note that

$$\|f\|_2^2 = \left\langle \sum_S a_S w_S, \sum_T a_T w_T \right\rangle = \sum_{S,T} a_S a_T \langle w_S, w_T \rangle = \sum_S a_S^2.$$

This is the so-called Parseval's identity. Recall that $f_i(x) = f(x) - f(x^i)$. It is easy to check that

$$\hat{f}_i(S) = \begin{cases} 0 & i \notin S \\ 2\hat{f}(S) & i \in S \end{cases}.$$

Therefore

$$\|f_i\|_2^2 = 4 \sum_{S\colon i \in S} a_S^2.$$

On the other hand,

$$|f_i(x)| = \begin{cases} 0 & f(x) = f(x^i) \\ 2 & f(x) \neq f(x^i) \end{cases}.$$

Thus

$$\|f_i\|_p^p = 2^p \mathbb{P}(f(x) \neq f(x^i)) = 2^p I_i(f).$$

Taking $p = 2$ we obtain

$$I_i(f) = \sum_{S\colon i \in S} a_S^2,$$

hence we have a crucial identity

$$I(f) = \sum_{i=1}^{n} \sum_{S\colon i \in S} a_S^2 = \sum_S |S| a_S^2.$$

connecting the total influence with the spectrum of f.

Let us define

$$\mathrm{Var}_\mu(f) = \mathbb{E}_\mu f^2 - (\mathbb{E}_\mu f)^2.$$

Note that we have

$$\mathbb{E} f = \sum_S a_S \mathbb{E} w_S = a_\emptyset.$$

Therefore

$$\mathrm{Var}_\mu(f) = \sum_S a_S^2 - a_\emptyset^2 = \sum_{S\colon |S| \geq 1} a_S^2.$$

On the other hand we have

$$\begin{aligned} \mathrm{Var}_\mu(f) &= \mathbb{E}_\mu f^2 - \big(\mathbb{E}_\mu f\big)^2 = 1 - \big(\mathbb{P}(f = 1) - \mathbb{P}(f = -1)\big)^2 \\ &= 1 - (2\mu(f) - 1)^2 = 4\mu(f)(1 - \mu(f)). \end{aligned}$$

Having this facts we can give a simple proof of the that Dict_n has the smallest influence among all functions with mean 0 (or, in other words, with $\mu(f) = 1/2$). Namely, we have

Proposition 3. Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $\mu(f) = \mathbb{P}(f = 1)$. Then we have

$$I(f) \geq 4\mu(f)(1 - \mu(f)).$$

In particular, if $\mu(f) = 1/2$ we obtain $I(f) \geq 1$.

Proof. The inequality is equivalent to $I(f) \geq \mathrm{Var}_\mu(f)$. This is true since

$$\mathrm{Var}_\mu(f) = \sum_{S\colon |S| \geq 1} a_S^2 \leq \sum_{S\colon |S| \geq 1} |S| a_S^2 = \sum_S |S| a_S^2 = I(f).$$

□

6 Hypercontractivity

The cube $\{-1,1\}^n$ possess a group structure. Namely, we can define the group multiplication by

$$(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = (x_1 y_1, \ldots x_n, y_n).$$

The measure μ is a Haar measure on $(\{-1,1\}, \cdot)$, i.e. $\mu(g \cdot A) = \mu(A)$ where $g \in \{-1,1\}^n$ and $A \subset \{-1,1\}^n$. Here $g \cdot A = \{g \cdot a : \ a \in A\}$.

Let ν be any a measure on $\{-1,1\}^n$. We define a convolution operator T_ν by the formula

$$T_\nu(f)(x) = \int f(xy^{-1}) \, \mathrm{d}\nu(y).$$

Since $y^{-1} = y$, we can write as well

$$T_\nu(f)(x) = \int f(xy) \, \mathrm{d}\nu(y).$$

This operator is a weak contraction in every $L_p(\{-1,1\}, \mu)$ for $p \geq 1$. Indeed, by triangle inequality an Jensens inequality we have

$$\begin{aligned} \|T_\nu(f)\|_p^p &= \int \left| \int f(xy^{-1}) \, \mathrm{d}\nu(y) \right|^p \mathrm{d}\mu(x) \leq \int \int \left| f(xy^{-1}) \right|^p \mathrm{d}\nu(y) \, \mathrm{d}\mu(x) \\ &= \int \int \left| f(xy^{-1}) \right|^p \mathrm{d}\mu(x) \, \mathrm{d}\nu(y) = \int \int |f(x)|^p \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y) = \|f\|_p^p. \end{aligned}$$

We have used the fact that μ is Haar measure on $\{-1,1\}^n$.

Now take

$$\nu_\delta^n = \left(\frac{1+\delta}{2} \delta_{\{1\}} + \frac{1-\delta}{2} \delta_{\{-1\}} \right)^{\otimes n}$$

and let $T_\delta = T_\delta^{(n)} = T_{\nu_\delta^n}$. We investigate the action of T_δ on Walsh functions,

$$\begin{aligned} T_\delta(w_S)(x) &= \int \prod_{i \in S} x_i y_i \, \mathrm{d}\nu_\delta^n(y) = \left(\prod_{i \in S} x_i \right) \left(\prod_{i \in S} \int y_i \, \mathrm{d}\nu_\delta(y_i) \right) \\ &= w_S(x) \delta^{|S|}. \end{aligned}$$

Therefore, if $f : \{-1,1\}^n \to \mathbb{R}$ then we have

$$T_\delta(f) = \sum_{S \subset [n]} a_s \delta^{|S|} w_S, \qquad \text{when} \quad f = \sum_{S \subset [n]} a_S w_S.$$

The operator T_δ possess the following properties

- $T_\delta(f) \geq f$, when $f \geq 0$,
- $T_\delta(1) = 1$,
- $\langle f, T_\delta g \rangle = \langle T_\delta f, g \rangle$,
- $\|T_\delta f\|_p \leq \|f\|_p$.

We are going to develop one of the most important tools in the theory of boolean functions, namely prove that T_δ is hypercontractive.

Theorem 1 (Bonami-Beckner-Gross)**.** For any $f : \{-1,1\}^n \to \mathbb{R}$ and any $\delta \in [0,1]$ we have

$$\|T_\delta f\|_2 \leq \|f\|_{1+\delta_2}.$$

We begin with the following abstract lemma.

Lemma 6. Let $q \geq p \geq 1$ and let (Ω_1, μ_1), (Ω_2, μ_2) be two finite probability spaces. Let $K_i : \Omega_i \times \Omega_i \to \mathbb{R}$ for $i = 1,2$. We define two operators

$$T_i(f)(x) = \int_{\Omega_i} K_i(x,y) \, \mathrm{d}\mu_i(y), \qquad i = 1,2.$$

Moreover, for $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ let us take

$$(T_1 \otimes T_2)(f)(x_1,x_2) = \int_{\Omega_1} \int_{\Omega_2} f(y_1,y_2) K_1(x_1,y_1) K_2(x_2,y_2) \, \mathrm{d}\mu_2(y_2) \, \mathrm{d}\mu_1(y_1).$$

Suppose that for $i = 1, 2$ we have

$$\|T_i f\|_{L_q(\Omega_i,\mu_i)} \leq \|f\|_{L_p(\Omega_i,\mu_i)}, \qquad \text{for all} \quad f : \Omega_i \to \mathbb{R}.$$

Then

$$\|T_1 \otimes T_2 f\|_{L_q(\Omega_1\times\Omega_2,\mu_1\otimes\mu_2)} \leq \|f\|_{L_p(\Omega_1\times\Omega_2,\mu_1\otimes\mu_2)}.$$

Proof. Take $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$. The operator T_2 acts on a functions $f : \Omega_1 \to \mathbb{R}$. However, we can define its action on functions of two variables by the formula

$$T_2(f)(y_1, x_2) = \int f(y_1, y_2) K_2(x_2, y_2) \, \mathrm{d}\mu_2(y_2).$$

Now it $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ then we have

$$T_1 \otimes T_2 f = T_1(T_2(f)).$$

More precisely,

$$(T_1 \otimes T_2)(f)(x_1, x_2) = T_1 \left(T_2(f)(\cdot, x_2)\right)(x_1).$$

By the assumption on T_1 we have

$$\begin{aligned}\|T_1 \otimes T_2 f\|^q_{L_q(\Omega_1\times\Omega_2,\mu_1\otimes\mu_2)} &= \int_{\Omega_2}\int_{\Omega_1} |T_1 \left(T_2(f)(\cdot, x_2)\right)(x_1)|^q \, \mathrm{d}\mu_1(x_1) \, \mathrm{d}\mu_2(x_2) \\ &\leq \int_{\Omega_2} \left(\int_{\Omega_1} |(T_2(f)(y_1, x_2))|^p \, \mathrm{d}\mu_1(y_1)\right)^{q/p} \mathrm{d}\mu_2(x_2).\end{aligned}$$

Now it (X, μ), (Y, ν) are finite probability spaces and $r \geq 1$ then we have the following Minkowski inequality

$$\left(\int_X \left(\int_Y g(x, y) \, \mathrm{d}\nu(y)\right)^r \mathrm{d}\mu(x)\right)^{1/r} \leq \int_Y \left(\int_X g(x, y)^r \, \mathrm{d}\mu(x)\right)^{1/r} \mathrm{d}\nu(y).$$

If we realize that the integral over Y in the above inequality is simply a finite sums then we shall see that this inequality means that

$$\|\sum_i a_i g_i\|_r \leq \sum_i a_i \|g_i\|_r,$$

where $g_i : X \to \mathbb{R}$ and (a_i) are positive numbers. This in is the usual well known Minkowski inequality.

We apply this inequality to the function

$$g(y_1, x_2) = |(T_2(f)(y_1, x_2))|^p$$

and $(X, \mu) = (\Omega_2, \mu_2)$, $(Y, \nu) = (\Omega_1, \mu_1)$, $r = q/p$,

$$\begin{aligned}\left(\int_{\Omega_2} \left(\int_{\Omega_1} |(T_2(f)(y_1, x_2))|^p \, \mathrm{d}\mu_1(y_1)\right)^{q/p} \mathrm{d}\mu_2(x_2)\right)^{p/q} \\ \le \left(\int_{\Omega_1} \left(\int_{\Omega_2} |(T_2(f)(y_1, x_2))|^q \, \mathrm{d}\mu_2(x_2)\right)^{p/q} \mathrm{d}\mu_1(y_1)\right).\end{aligned}$$

It follow that

$$\begin{aligned}\int_{\Omega_2} \left(\int_{\Omega_1} |(T_2(f)(y_1, x_2))|^p \, \mathrm{d}\mu_1(y_1)\right)^{q/p} \mathrm{d}\mu_2(x_2) \\ \le \left(\int_{\Omega_1} \left(\int_{\Omega_2} |(T_2(f)(y_1, x_2))|^q \, \mathrm{d}\mu_2(x_2)\right)^{p/q} \mathrm{d}\mu_1(y_1)\right)^{q/p}.\end{aligned}$$

Now we apply our assumption on T_2 and obtain

$$\left(\int_{\Omega_2} |(T_2(f)(y_1, x_2))|^q \, \mathrm{d}\mu_2(x_2)\right)^{1/q} \le \left(\int_{\Omega_2} |f(y_1, y_2)|^p \, \mathrm{d}\mu_2(y_2)\right)^{1/p}.$$

Thus,

$$\begin{aligned}\int_{\Omega_1} \left(\int_{\Omega_2} |(T_2(f)(y_1, x_2))|^q \, \mathrm{d}\mu_2(x_2)\right)^{p/q} \mathrm{d}\mu_1(y_1) \\ \le \int_{\Omega_1} \int_{\Omega_2} |f(y_1, y_2)|^p \, \mathrm{d}\mu_2(y_2) \, \mathrm{d}\mu_1(y_1)\end{aligned}$$

We arrive at

$$\begin{aligned}\|T_1 \otimes T_2 f\|^q_{L_q(\Omega_1 \times \Omega_2, \mu_1 \otimes \mu_2)} &\le \left(\int_{\Omega_1} \int_{\Omega_2} |f(y_1, y_2)|^p \, \mathrm{d}\mu_2(y_2) \, \mathrm{d}\mu_1(y_1)\right)^{q/p} \\ &= \|f\|^q_{L_p(\Omega_1 \times \Omega_2, \mu_1 \otimes \mu_2)}.\end{aligned}$$

$\square$

Note that in the case $n = 1$ we have

$$T_\delta^{(1)}(f)(x) = \frac{1+\delta}{2} f(x) + \frac{1-\delta}{2} f(-x) = \int_{\{-1,\}1} f(xy)(1+\delta y)\,\mathrm{d}\mu(y)$$
$$= \int_{\{-1,\}1} f(y)(1+\delta y x^{-1})\,\mathrm{d}\mu(y).$$

In general,

$$\begin{aligned} T_\delta^{(n)}(f)(x) &= \int_{\{-1,1\}^n} f(x_1y_1, \ldots, x_ny_n)\,\mathrm{d}\nu_\delta^{(1)}(y_1)...\mathrm{d}\nu_\delta^{(1)}(y_n) \\ &= \int_{\{-1,1\}^n} f(y_1, \ldots, y_n)(1+\delta y_1 x_1^{-1}) \ldots (1+\delta y_n x_n^{-1})\,\mathrm{d}\mu^{(1)}(y_1)...\mathrm{d}\mu^{(1)}(y_n) \\ &= \int_{\{-1,1\}^n} f(y_1, \ldots, y_n) K(x_1, y_1) \ldots K(x_n, y_n)\,\mathrm{d}\mu^{(1)}(y_1)...\mathrm{d}\mu^{(1)}(y_n), \end{aligned}$$

where

$$K(x, y) = 1 + \delta y x^{-1}.$$

Therefore, using induction and Lemma 6 we reduce the proof of the Theorem 1 to the case $n = 1$. In this case we have

$$(T_\delta f)(x) = \frac{1+\delta}{2} f(x) + \frac{1-\delta}{2} f(-x).$$

Therefore,

$$\|T_\delta f\|_2 = \left(\frac{\left|\frac{1+\delta}{2} f(1) + \frac{1-\delta}{2} f(-1)\right|^2 + \left|\frac{1+\delta}{2} f(-1) + \frac{1-\delta}{2} f(1)\right|^2}{2} \right)^{1/2}$$

and

$$\|f\|_{1+\delta^2} = \left(\frac{|f(1)|^{1+\delta^2} + |f(-1)|^{1+\delta^2}}{2} \right)^{\frac{1}{1+\delta^2}}.$$

Let

$$a = \frac{f(1) + f(-1)}{2}, \qquad b = \frac{f(1) - f(-1)}{2}.$$

The inequality $\|T_\delta f\|_2 \le \|f\|_{1+\delta^2}$ is now equivalent to

$$\left(\frac{|a + b\delta|^2 + |a - b\delta|^2}{2} \right)^{1/2} \le \left(\frac{|a + b|^{1+\delta} + |a - b|^{1+\delta^2}}{2} \right)^{\frac{1}{1+\delta^2}}.$$

Since
$$\frac{|a+b\delta|^2+|a-b\delta|^2}{2}=a^2+\delta^2b^2,$$
we have to prove the following lemma.

Lemma 7. For all $a,b\in\mathbb{R}$ and $\delta\in[0,1]$ we have an inequality
$$(a^2+b^2\delta^2)^{\frac{1+\delta^2}{2}}\le\frac{|a+b|^{1+\delta^2}+|a-b|^{1+\delta^2}}{2}.$$

Proof. If $a=0$ then our inequality has the form $|b|^{1+\delta^2}\delta^{1+\delta^2}\le|b|^{1+\delta^2}$, which is true since $\delta^{1+\delta^2}\le 1^{1+\delta^2}=1$. Therefore we can assume that $a\ne 0$. If we divide both sides of the inequality by $|a|^{1+\delta^2}$ and denote $y=b/a$ we are to prove
$$(1+\delta^2y^2)^{\frac{1+\delta^2}{2}}\le\frac{|1+y|^{1+\delta^2}+|1-y|^{1+\delta^2}}{2}.$$
Both sides of this inequality are even functions of the variable y. Therefore one can assume that $y\ge 0$.

Let us first consider the case $y\in[0,1)$. We have the following Taylor expansion
$$(1+x)^{1+\delta^2}=\sum_{k=0}^{\infty}\binom{1+\delta^2}{k}x^k,\qquad |x|<1,$$
where
$$\binom{1+\delta^2}{k}=\frac{(1+\delta^2)(1+\delta^2-1)\dots(1+\delta^2-k+1)}{k!}.$$
Thus,
$$\begin{aligned}\frac{|1+y|^{1+\delta^2}+|1-y|^{1+\delta^2}}{2}&=\frac{1}{2}\left[\sum_{k=0}^{\infty}\binom{1+\delta^2}{k}y^k+\sum_{k=0}^{\infty}\binom{1+\delta^2}{k}(-y)^k\right]\\&=\sum_{k=0}^{\infty}\binom{1+\delta^2}{2k}y^{2k}=1+\frac{(1+\delta^2)\delta^2}{2}y^2+\sum_{k=2}^{\infty}\binom{1+\delta^2}{2k}y^{2k}\\&\ge 1+\frac{(1+\delta^2)\delta^2}{2}y^2,\end{aligned}$$
since
$$\binom{1+\delta^2}{2k}=\frac{(1+\delta^2)(1+\delta^2-1)\dots(1+\delta^2-2k+1)}{(2k)!}\ge 0$$

as in the numerator there are 2 positive term and $2k$ negative terms. It suffices to prove

$$(1+\delta^2 y^2)^{\frac{1+\delta^2}{2}} \le 1 + \frac{(1+\delta^2)\delta^2}{2} y^2. \tag{1}$$

Note that $(1+x)^\lambda \le 1+\lambda x$ for $x \ge 0$ and $\lambda \in [0,1]$. This is called the Bernoulli inequality. It follows from the fact that $g(x) = (1+x)^\lambda - 1 - \lambda x$ satisfies $g(0) = 0$ and $g'(x) \le 0$ for $x \ge 0$. Taking $x = \delta^2 y^2$ and $\lambda = \frac{1+\delta^2}{2}$ we obtain (1).

The case $y = 1$ follows from the previous case by continuity.

Let us now consider the case $y > 1$. Take $z = \frac{1}{y} < 1$. We are to prove that

$$\left(1 + \frac{\delta^2}{z^2}\right)^{\frac{1+\delta^2}{2}} \le \frac{\left|1+\frac{1}{z}\right|^{1+\delta^2} + \left|1-\frac{1}{z}\right|^{1+\delta^2}}{2}.$$

Multiplying both sides by $z^{1+\delta^2}$ we obtain

$$\left(z^2+\delta^2\right)^{\frac{1+\delta^2}{2}} \le \frac{|1+z|^{1+\delta^2} + |1-z|^{1+\delta^2}}{2}.$$

This follows from the first case, since

$$z^2 + \delta^2 = 1 + \delta^2 z^2 - (1-z^2)(1-\delta^2) \le 1 + \delta^2 z^2.$$

□

7 KKL Theorem and Talagrand's theorem

We are now ready to prove the following celebrated KKL Theorem.

Theorem 2 (Kahn-Kalai-Linial)**.** Suppose $f : \{-1,1\}^n \to \{-1,1\}$ with $\mu(f) = p \le \frac{1}{2}$. Then

$$\sum_{i=1}^n I_i(f)^2 \ge C^2 p^2 \frac{(\ln n)^2}{n}.$$

Moreover,

$$\max_{1 \le i \le n} I_i(f) \ge Cp \frac{\ln n}{n}.$$

Proof. Since
$$\sum_{i=1}^{n} I_i(f)^2 \le n \left(\max_{1\le i\le n} I_i(f) \right)^2,$$
the second inequality follows directly from the first one.

Let $f_i(x) = f(x) - f(x^i) \in \{-2, 0, 2\}$. Hypercontractivity yields
$$\|T_\delta f_i\|_2 \le \|f_i\|_{1+\delta^2}, \qquad \delta \in [0,1].$$
Recall that
$$\hat{f}_i(S) = \begin{cases} 0 & i \notin S \\ 2\hat{f}(S) & i \in S \end{cases}.$$
Therefore, if $f = \sum a_S w_S$ then
$$f_i = 2 \sum_{S\colon\, i\in S} a_s w_S$$
and
$$\|f_i\|_2^2 = 4 \sum_{S\colon\, i\in S} a_S^2.$$
Moreover,
$$T_\delta f_i = 2 \sum_{i\colon\, i\in S} a_S \delta^{|S|} w_S$$
and
$$\|T_\delta f_i\|_2^2 = 4 \sum_{S\colon\, i\in S} a_S^2 \delta^{2|S|}.$$
On the other hand, for $p \ge 1$ we have
$$\|f_i\|_p^p = 2^p \mathbb{P}(f(x) \ne f(x^i)) = 2^p I_i,$$
where $I_i = I_i(f)$. Thus,
$$4 \sum_{S\colon\, i\in S} a_S^2 \delta^{2|S|} \le \|f_i\|_{1+\delta^2}^2 = \left(\|f_i\|_{1+\delta^2}^{1+\delta^2} \right)^{\frac{2}{1+\delta^2}} = \left(2^{1+\delta^2} I_i \right)^{\frac{2}{1+\delta^2}} = 4 I_i^{\frac{2}{1+\delta^2}}.$$
Summing these inequalities for $1 \le i \le n$ we obtain
$$\sum_S a_S^2 |S| \delta^{2|S|} \le \sum_{i=1}^{n} I_i^{\frac{2}{1+\delta^2}}.$$

Hence,

$$\delta^{2|S|} \sum_{S\colon\ |S|\le M} a_S^2|S| \le \sum_{S\colon\ |S|\le M} a_S^2|S|\delta^{2|S|} \le \sum_S a_S^2|S|\delta^{2|S|} \le \sum_{i=1}^n I_i^{\frac{2}{1+\delta^2}}.$$

We have

$$\sum_S a_S^2 = 1, \quad a_\emptyset = p-(1-p) = 2p-1.$$

Note that

$$\sum_{S\colon\ |S|\le M} a_S^2|S| \ge \sum_{S\colon\ |S|\le M} a_S^2 - a_\emptyset^2.$$

Therefore,

$$\delta^{-2M}\sum_{i=1}^n I_i^{\frac{2}{1+\delta^2}} \ge \sum_{S\colon\ |S|\le M} a_S^2 - a_\emptyset^2.$$

Since

$$\sum_{i=1}^n I_i = \sum_S |S|a_S^2,$$

then we also have

$$\sum_{i=1}^n I_i \ge M\sum_{|S|>M} a_S^2.$$

Summing these two inequalities we obtain

$$\sum_{i=1}^n \left(\delta^{-2M} I_i^{\frac{2}{1+\delta^2}} + \frac{1}{M}I_i\right) \ge \sum_S a_S^2 - a_\emptyset^2 = 1-(2p-1)^2 = 4p(1-p) \ge 2p.$$

Let $\lambda \ge 0$ be a number satisfying $\sum_{i=1}^n I_i^2 = \frac{\lambda^2}{n}$. Suppose, by contradiction, that $\lambda < Cp\ln n$. We show that for small values of C this in impossible.

We have

$$\sum_{i=1}^n I_i \le \sqrt{n}\sqrt{\sum_{i=1}^n I_i^2} = \lambda.$$

Moreover, by Jensen inequality we have

$$\sum_{i=1}^n I_i^{\frac{2}{1+\delta^2}} \le n\left(\frac{1}{n}\sum_{i=1}^n \left(I_i^{\frac{2}{1+\delta^2}}\right)^{\frac{1}{1+\delta^2}}\right)^{\frac{1}{1+\delta^2}} = n\left(\frac{\lambda^2}{n^2}\right)^{\frac{1}{1+\delta^2}}$$
$$= \lambda^{\frac{2}{1+\delta^2}} n^{1-\frac{2}{1+\delta^2}} = \lambda^{\frac{2}{1+\delta^2}} n^{\frac{\delta^2-1}{\delta^2+1}}.$$

Thus,

$$2p \leq \sum_{i=1}^{n} \left(\delta^{-2M} I_i^{\frac{2}{1+\delta^2}} + \frac{1}{M} I_i \right) \leq \delta^{-2M} \lambda^{\frac{2}{1+\delta^2}} n^{\frac{\delta^2-1}{\delta^2+1}} + \frac{\lambda}{M}.$$

Let $M = \lceil \lambda/p \rceil$. Then

$$\frac{\lambda}{p} \leq M \leq 1 + \frac{\lambda}{p} \leq 1 + C \ln n.$$

Thus,

$$2p \leq \delta^{-2M} \lambda^{\frac{2}{1+\delta^2}} n^{\frac{\delta^2-1}{\delta^2+1}} + \frac{\lambda}{M} \leq \delta^{-2(1+C\ln n)} (Cp \ln n)^{\frac{2}{1+\delta^2}} n^{\frac{\delta^2-1}{\delta^2+1}} + p.$$

This is equivalent to

$$1 \leq p^{\frac{1-\delta^2}{1+\delta^2}} \delta^{-2(1+C\ln n)} (Cp \ln n)^{\frac{2}{1+\delta^2}} n^{\frac{\delta^2-1}{\delta^2+1}}.$$

Taking $\delta = 1/2$ and using $p \leq 1/2$ we obtain

$$1 \leq \left(\frac{1}{2}\right)^{3/5} 2^{2(1+C\ln n)} C^{8/5} n^{-\frac{3}{5}} (\ln n)^{8/5} = 2^{7/5} C^{8/5} n^{-\frac{3}{5}+2C\ln 2} (\ln n)^{8/5}.$$

Take $C < \frac{1}{5\ln 2}$. Then

$$1 \leq 2^{7/5} C^{8/5} n^{-\frac{1}{5}} (\ln n)^{8/5} \leq C^{8/5} c_0,$$

where c_0 is an universal constant. Now it suffices to take sufficiently small C to obtain a contradiction. □

We prove another theorem of this kind (due to Talagrand) and show that KKL Theorem follows from this theorem.

Theorem 3. Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $\mu(f) = \mathbb{P}(f = 1)$. Then

$$\sum_{i=1}^{n} \frac{I_i(f)}{\log\left(\frac{1}{I_i(f)}\right)} \geq \frac{4}{15} \mu(f)(1-\mu(f)).$$

We adopt the notation $\frac{0}{\log(1/0)} = 0$ and $1/\log(1) = +\infty$. We begin with a lemma.

Lemma 8. Let $g : \{-1,1\}^n \to \mathbb{R}$ with $\|g\|_{3/2} \neq \|g\|_2$, which is equivalent to $|g|$ being not constant. Then

$$\sum_{S \neq \emptyset} \frac{\hat{g}(S)^2}{|S|} \leq \frac{5}{2} \frac{\|g\|_2^2}{\log\left(\|g\|_2/\|g\|_{3/2}\right)}.$$

Proof. Using the inequality

$$\|T_\delta g\|_2 \le \|g\|_{1+\delta^2}$$

with $\delta^2 = 1/2$ we obtain

$$\sum_{S:\ |S|=k} \hat{g}(S)^2 \le 2^k \sum_S \frac{1}{2^{|S|}} \hat{g}(S)^2 = 2^k \|T_{\sqrt{1/2}} g\|_2^2 \le 2^k \|g\|_{3/2}^2.$$

Now take $m \ge 0$. We have

$$\begin{aligned} \sum_{S \ne \emptyset} \frac{\hat{g}(S)^2}{|S|} &= \sum_{k=1}^m \sum_{S:\ |S|=k} \frac{\hat{g}(S)^2}{k} + \sum_{S:\ |S|>m} \frac{\hat{g}(S)^2}{|S|} \le \sum_{k=1}^m \frac{2^k \|g\|_{3/2}^2}{k} + \sum_{S:\ |S|>m} \frac{\hat{g}(S)^2}{m+1} \\ &\le \frac{4 \cdot 2^m \|g\|_{3/2}^2 + \|g\|_2^2}{m+1}, \end{aligned}$$

where we have used the inequality

$$\sum_{k=1}^m \frac{2^k}{k} \le \frac{4 \cdot 2^m}{m+1},$$

which can be easily proved by induction.

Now we take

$$m = \max\{m \ge 0 \mid 2^m \|g\|_{3/2}^2 \le \|g\|_2^2\}.$$

Then $2^{m+1}\|g\|_{3/2}^2 > \|g\|_2^2$. Hence,

$$m + 1 > 2 \log \left(\frac{\|g\|_2}{\|g\|_{3/2}} \right).$$

We arrive at

$$\sum_{S \ne \emptyset} \frac{\hat{g}(S)^2}{|S|} \le \frac{5\|g\|_2^2}{m+1} \le \frac{5}{2} \frac{\|g\|_2^2}{\log \left(\|g\|_2 / \|g\|_{3/2} \right)}.$$

□

Proof of Talagrand's theorem. Suppose $I_i(f) \in (0,1)$. Let $g(x) = f(x) - f(x^i)$. It follows that $|g|$ is not constant. We have

$$\frac{\|g\|_2}{\|g\|_{3/2}} = \frac{2I_i(f)^{1/2}}{2I_i(f)^{2/3}} = I_i(f)^{-1/6}.$$

From the lemma we obtain

$$\sum_{S\colon\, i\in S}\frac{4\hat{f}(S)^2}{|S|}=\sum_{S}\frac{\hat{g}(S)^2}{|S|}\le\frac{5}{2}\frac{\|g\|_2^2}{\log\left(\|g\|_2/\|g\|_{3/2}\right)}=\frac{5}{2}\cdot\frac{4I_i(f)}{\log(I_i(f)^{-1/6})}=60\frac{I_i(f)}{\log(\frac{1}{I_i(f)})}.$$

The inequality

$$\sum_{S\colon\, i\in S}\frac{4\hat{f}(S)^2}{|S|}\le 60\frac{I_i(f)}{\log(\frac{1}{I_i(f)})}$$

is also true when $I_i(f)\in\{0,1\}$. We obtain

$$16\mu(f)(1-\mu(f))=4\operatorname{Var}_\mu(f)=\sum_{Sn\varepsilon\emptyset}4\hat{f}(S)^2=\sum_{i=1}^{n}\sum_{S\colon\, i\in S}\frac{4\hat{f}(S)^2}{|S|}\le 60\sum_{i=1}^{n}\frac{I_i(f)}{\log(\frac{1}{I_i(f)})}.$$

The assertion follows. □

We show that Talagrand result implies KKL Theorem. Let us first observe that if $a\in(0,1)$ and $\frac{a}{\log(1/a)}\ge c>0$ then $a\ge\frac{1}{2}c\log(1/c)$. Since $(0,1)\ni a\mapsto\frac{a}{\log(1/a)}$ is increasing, it suffices to assume that $\frac{a}{\log(1/a)}=c$. Then we are to prove

$$a\ge\frac{1}{2}\frac{a}{\log(1/a)}\log\left(\frac{1}{a}\log\left(\frac{1}{a}\right)\right).$$

Taking $x=1/a\ge 1$ we see that this inequality is equivalent to

$$\log(x)\ge\frac{1}{2}\log(x\log(x))=\frac{1}{2}\log x+\frac{1}{2}\log\log x.$$

Thus we are to prove $x\ge\log x$. It follows from Bernoulli inequality

$$2^x=(1+1)^x\ge 1+x\ge x.$$

From Talagrand's inequality we know that there exists i such that

$$\frac{I_i(f)}{\log\left(\frac{1}{I_i(f)}\right)}\ge\frac{1}{n}\cdot\frac{4}{15}\mu(f)(1-\mu(f)).$$

Now take

$$a=I_i(f),\qquad c=\frac{1}{n}\cdot\frac{4}{15}\mu(f)(1-\mu(f)).$$

We have

$$\frac{1}{c}=n\cdot\frac{15}{4}\frac{1}{\mu(f)(1-\mu(f))}\ge 15n.$$

We obtain

$$I_i(f)\ge\frac{1}{2}c\log(1/c)\ge\frac{1}{n}\cdot\frac{4}{15}\mu(f)(1-\mu(f))\log(15n)\ge\frac{4}{15}\mu(f)(1-\mu(f))\frac{\log n}{n}.$$

This is the KKL Theorem.

8 Monotone boolean functions

The function $f : \{-1,1\}^n \to \{-1,1\}$ is called monotone if $x_i \le y_i$ for $1 \le i \le n$ implies $f(x_1,\dots,x_n) \le f(y_1,\dots,y_n)$. We calculate the influence of a monotone function f. Note that

$$\hat{f}(\{1\}) = \mathbb{E}x_1 f = \frac{1}{2}\mathbb{E}f(1,x_2,\dots x_n) - \frac{1}{2}\mathbb{E}f(-1,x_2,\dots,x_n).$$

Since our function is monotone, the difference

$$f(1,x_2,\dots x_n) - f(-1,x_2,\dots,x_n)$$

can have only values 0 and 2. Therefore,

$$\begin{aligned}\hat{f}(\{1\}) &= \frac{1}{2}\mathbb{E}(f(1,x_2,\dots x_n) - f(-1,x_2,\dots,x_n)) \\ &= \frac{1}{2}\cdot 2\mathbb{P}(f(1,x_2,\dots x_n) \neq f(-1,x_2,\dots,x_n)) = I_1(f).\end{aligned}$$

Therefore, for a monotone boolean function we have

$$I_i(f) = \hat{f}(\{i\}), \quad 1 \le i \le n, \qquad I(f) = \sum_{i=1}^{n} \hat{f}(\{i\}).$$

For an arbitrary boolean function f we can write

$$\begin{aligned}|a_i| &= \frac{1}{2}\left|\mathbb{E}f(1,x_2,\dots x_n) - f(-1,x_2,\dots,x_n)\right| \\ &\le \frac{1}{2}\mathbb{E}\left|f(1,x_2,\dots x_n) - f(-1,x_2,\dots,x_n)\right| \\ &= \mathbb{P}(f(1,x_2,\dots x_n) \neq f(-1,x_2,\dots,x_n)).\end{aligned}$$

Thus

$$|a_i| \le I_i(f).$$

We can now easily prove the following estimate.

Proposition 4. Let $f : \{-1,1\}^n \to \{-1,1\}$ be a monotone boolean function. Then

$$I(f) \le \sqrt{n}$$

Proof. We have

$$I(f)=\sum_{i=1}^n \hat{f}(\{i\}) \le \sqrt{n}\sum_{i=1}^n \hat{f}(\{i\})^2 \le \sqrt{n}\sum_S \hat{f}(S)^2=\sqrt{n}.$$

$\square$

Now we introduce certain symmetrization techniques. Namely we prove the following proposition.

Proposition 5. Let $f:\{-1,1\}\to\{-1,1\}$. Then there exists $g:\{-1,1\}\to\{-1,1\}$ such that $\mathbb{E}f=\mathbb{E}g$ and $I_i(f)\ge I_i(g)$.

Proof. For $1\le i\le n$ we take the ith symmetrization of f given by the formula

$$f_{s_i}(x_1,...,x_n)=\begin{cases} f(x_1,...,x_n) & f(...,x_{i-1},-1,x_{i+1},...)\le f(...,x_{i-1},-1,x_{i+1},...) \\ -f(x_1,...,x_n) & f(...,x_{i-1},-1,x_{i+1},...)> f(...,x_{i-1},-1,x_{i+1},...) \end{cases}$$

Clearly $I_i(f)=I_i(f_{s_i})$. To check that $I_j(f)\ge I_j(f_{s_i})$ for $i\ne j$ it suffices to consider $i=1, j=2$. Now one has to consider elements

$$(-1,-1,x),(-1,1,x),(1,-1,x),(1,1,x)\in\{-1,1\}^n$$

and 16 possible values of f in these points. It suffices to observe that the contribution to I_2 will change only when

$$f(-1,-1,x)\ne f(-1,1) \quad \text{and} \quad f(1,-1,x)\ne f(1,1,x)$$

and I_2 will decrease.

Now, we construct a sequence of symmetrizations $f, f_{s_{i_1}}, f_{s_1,s_2}=(f_{s_{i_1}})_{s_{i_2}},\ldots$ in the following way: whenever we have a function $f_{s_1,...,s_k}$ that is not monotone we find a direction s_{k+1} for which we can do non-trivial symmetrization and then we take $f_{s_1,...,s_k,s_{k+1}}$. We only have to show that this procedure will stop. But this is clear since the functional

$$\mathcal{L}(f)=\sum_{x\in\{-1,1\}^n}(1+f(x))(x_1+\ldots+x_n)$$

satisfies $\mathcal{L}(f)<\mathcal{L}(f_{s_i})$ and $\mathcal{L}(f)\le 2n2^{n-1}$. $\square$

Take $p\in[0,1]$ and let

$$\mu_p=\left((1-p)\delta_{\{-1\}}+p\delta_{\{1\}}\right)^{\otimes n}$$

and let $\mu_p(f)=\mu_p(\{f=1\})$. Moreover, let $I_i^p(f)=\mu_p(f(x)\ne f(x^i))$ and $I^p(f)=\sum_{i=1}^n I_i^p(f)$. We prove the following famous Margulis-Russo lemma.

Lemma 9 (Margulis-Russo lemma)**.** Let $f : \{-1, 1\} \to \{-1, 1\}$ be monotone. Then we have

$$\frac{\mathrm{d}}{\mathrm{d}p}\mu_p(f) = I^p(f).$$

Proof. Instead of μ_p let us consider

$$\mu_{p_1,\ldots,p_n} = \left((1-p_1)\delta_{\{-1\}} + p_1\delta_{\{1\}}\right) \otimes \ldots \otimes \left((1-p_n)\delta_{\{-1\}} + p_n\delta_{\{1\}}\right).$$

We claim that

$$\frac{\partial \mu_{p_1,\ldots,p_n}(f)}{\partial p_i} = I_i^{(p_1,\ldots,p_n)}(f).$$

Then by the chain role we have

$$\frac{\mathrm{d}\mu_p(f)}{\mathrm{d}p} = \sum_{i=1}^{n} \frac{\partial \mu_{p_1,\ldots,p_n}(f)}{\partial p_i}\Big|_{p_1=\ldots=p_n=p} = \sum_{i=1}^{n} I_i^{(p,\ldots,p)}(f) = \sum_{i=1}^{n} I_i^p(f).$$

Now we prove our claim. It suffices to take $i = 1$. Let $f_1(x) = f(x) - f(x^i)$. We have

$$\mathbb{P}_{p_1,\ldots,p_n}(f = 1) = \mathbb{P}_{p_1,\ldots,p_n}(f = 1, f_1 \neq 0) + \mathbb{P}_{p_1,\ldots,p_n}(f = 1, f_1 = 0).$$

Let $A \subset \{-1, 1\}^{n-1}$ be defined as follows,

$$A = \left\{x \in \{-1, 1\}^{n-1} \mid f(1, x) = 1, f_1(1, x) = 0\right\}.$$

If $f(1, x) = 1$ and $f_1(1, x) = 0$ then $f(-1, x) = 1$ and $f_1(-1, x) = 0$. Therefore

$$\{f = 0, f_i = 0\} = \{-1, 1\} \times A.$$

hence

$$\mathbb{P}_{p_1,\ldots,p_n}(f = 1, f_1 = 0) = \mathbb{P}_{p_2,\ldots,p_n}(A)$$

and therefore it does not depend on p_1.

Since f is monotone we have

$$\{f = 1, f_1 \neq 0\} = \{(x_1, \ldots, x_n) \mid x_1 = 1, f(1, \ldots, x_n) = 1, f(-1, \ldots, x_n) = -1, \}.$$

Define $B \subset \{-1, 1\}^{n-1}$ by

$$B = \left\{x \in \{-1, 1\}^{n-1} \mid f(1, x) = 1, f_1(1, x) \neq 0\right\}.$$

It follows that

$$\{f = 1, f_1 = 0\} = \{1\} \times B.$$

Therefore,

$$\mathbb{P}_{p_1,\dots,p_n}(f=1, f_1 \neq 0) = p_1 \mathbb{P}_{p_2,\dots,p_n}(B)$$

Note also that

$$I_1^{(p_1,\dots,p_n)}(f) = \mu_{p_1,\dots,p_n}\left(\{-1,1\} \times B\right) = \mathbb{P}_{p_2,\dots,p_n}(B).$$

Thus

$$\frac{\partial \mu_{p_1,\dots,p_n}(f)}{\partial p_1} = \frac{\partial}{\partial p_1}\left(\mathbb{P}_{p_2,\dots,p_n}(A) + p_1 \mathbb{P}_{p_2,\dots,p_n}(B)\right) = \mathbb{P}_{p_2,\dots,p_n}(B) = I_1^{(p_1,\dots,p_n)}(f).$$

□

Show that among all monotone Boolean functions Maj_n is the one with largest influence. Namely we have

Proposition 6. Let n be odd. Then for every monotone $f : \{-1,1\}^n \to \{-1,1\}$ we have

$$I(f) \leq I(\mathrm{Maj}_n).$$

Proof. We use Margulis-Russo lemma,

$$\begin{aligned} I^p(f) = \frac{\mathrm{d}\mu_p(f)}{\mathrm{d}p} &= \frac{\mathrm{d}}{\mathrm{d}p}\left(\sum_{x:f(x)=1} p^{|S|}(1-p)^{n-|S|} f(x)\right) \\ &= \sum_{x:f(x)=1} p^{|S|}(1-p)^{n-|S|}\left(\frac{|S|}{p} - \frac{n-|S|}{1-p}\right) f(x). \end{aligned}$$

Taking $p = \frac{1}{2}$ we obtain

$$I(f) = \frac{1}{2^{n-1}} \sum_{x:f(x)=1} (2|S| - n) f(x).$$

To maximize the right hand side one has to take

$$f(x) = \begin{cases} 1 & 2|S| - n \geq 0 \\ -1 & 2|S| - n < 0 \end{cases}.$$

Clearly, this function is Maj_n. □

9 Friedgut's Theorem

We begin this section with the following problem. Suppose we have a boolean function $f : \{-1,1\}^n \to \{-1,1\}$ and we have a fixed $J \subset [n]$. We would like to find the best approximation of f in the L_2 norm with a function depending only on variables x_j with $j \in J$.

Suppose we want our approximation g to be real valued. For $f : \{-1,1\}^n \to \{-1,1\}$ we write

$$f(x_1, x_2, \ldots, x_n) = f(x_J, x_{J'}),$$

where $x_J = (x_{j_1}, \ldots, x_{j_{|J|}})$ represents the part of the vector x with variables labelled by the numbers in subset J. The vector $x_{J'}$ represents the rest of variables. We have

$$\|f-g\|_2^2 = \frac{1}{2^n} \sum_{x_J, x_{J'}} \left(f(x_J, x_{J'}) - g(x_J)\right)^2 = \frac{1}{2^n} \sum_{x_J} \sum_{x_{J'}} \left(f(x_J, x_{J'}) - g(x_J)\right)^2 .$$

To minimize the expression

$$\sum_{x_{J'}} \left(f(x_J, x_{J'}) - g(x_J)\right)^2 .$$

One can easily see that, having a real numbers $a_1, \ldots, a_N$ fixed, the quadratic function

$$x \mapsto \sum_{i=1}^{N} (a_i - x)^2$$

has a minimum in a point

$$x = \frac{\sum_{i=1}^N a_i}{N}.$$

Therefore we take

$$g(x_J) = \frac{1}{2^{n-|J|}} \sum_{x_{J'}} f(x_J, x_{J'}).$$

In other words,

$$g(x_J) = \mathbb{E}(f|x_J).$$

Taking this function g we obtain

$$\begin{aligned}\|f-g\|_2^2 = \frac{1}{2^n}\sum_{x_J,x_{J'}}\left(f(x_J,x_{J'})-g(x_J)\right)^2 = \frac{1}{2^n}\sum_{x_J,x_{J'}} f(x_J,x_{J'})^2 - \\ \frac{1}{2^{n-1}}\sum_{x_J,x_{J'}} f(x_J,x_{J'})g(x_J) + \frac{1}{2^n}\sum_{x_J,x_{J'}} g(x_J)^2 \\ = 1 - \frac{1}{2^{n-1}}2^{n-|J|}\sum_{x_J} g(x_J)^2 + \frac{1}{2^n}2^{n-|J|}\sum_{x_J} g(x_J)^2 \\ = 1 - 2^{-|J|}\sum_{x_J} g(x_J)^2 = 2^{-|J|}\sum_{x_J}\left(1-g(x_J)^2\right) \\ = 2^{-|J|}\sum_{x_J}\left(1-g(x_J)\right)\left(1+g(x_J)\right).\end{aligned}$$

Let $p(x) = \mathbb{P}(f=1|x)$. Then

$$g(x) = \mathbb{E}(f|x) = p(x) - (1-p(x)) = 2p(x) - 1.$$

Thus

$$\|f-g\|_2^2 = 2^{-|J|}\sum_{x_J} 4p(x_J)(1-p(x_J)).$$

Now we would like to investigate the approximation with $\{-1,1\}$-valued functions. Recall that we have

$$\|f-g\|_2^2 = \frac{1}{2^n}\sum_{x_J}\sum_{x_{J'}}\left(f(x_J,x_{J'})-g(x_J)\right)^2.$$

We are to minimize the expression of the form

$$\{-1,1\}\ni x \mapsto \sum_{i=1}^{N}(a_i-x)^2,$$

where $a_1,\ldots,a_N \in \{-1,1\}$ are fixed. Let $k = |\{1\le i\le N:\ a_i = 1\}|$. Therefore

$$\sum_{i=1}^{N}(a_i-x)^2 = k(1-x)^2 + (n-k)(1+x)^2.$$

Therefore we should take $x=1$ if $n-k\ge k$ and $x=-1$ if $n-k<k$. Since

$$\frac{1}{2^{|J|}}|\{x_{J'}:\ f(x_J,x_{J'})=1\}| = \mathbb{P}(f=1|x_J),$$

we should take

$$g(x_J) = \begin{cases} 1 & \mathbb{P}(f = 1|x_J) \geq \frac{1}{2} \\ -1 & \mathbb{P}(f = -1|x_J) < \frac{1}{2} \end{cases}.$$

We arrive at

$$\begin{aligned}\|f - g\|_2^2 = \frac{1}{2^n} \sum_{x_J, x_{J'}} (f(x_J, x_{J'}) - g(x_J))^2 = \frac{1}{2^n} \sum_{x_J, x_{J'}} f(x_J, x_{J'})^2 - \\ \frac{1}{2^{n-1}} \sum_{x_J, x_{J'}} f(x_J, x_{J'}) g(x_J) + \frac{1}{2^n} \sum_{x_J, x_{J'}} g(x_J)^2 \\ = 2 - \frac{1}{2^{n-1}} \sum_{x_J, x_{J'}} f(x_J, x_{J'}) g(x_J).\end{aligned}$$

Now

$$\sum_{x_{J'}} f(x_J, x_{J'}) = 2^{n-|J|} \left(p(x_J) - (1 - p(x_J))\right) = 2^{n-|J|} \left(2p(x_J) - 1\right).$$

Therefore,

$$\|f - g\|_2^2 = 2 - \frac{1}{2^{n-1}} 2^{n-|J|} \sum_{x_J} (2p(x_J) - 1) g(x_J) = 2 \cdot 2^{-|J|} \sum_{x_J} (1 - (2p(x_J) - 1) g(x_J))$$

We have

$$\begin{aligned} 1 - (2p(x_J) - 1) g(x_J) &= \begin{cases} 1 - (2p(x_J - 1)) & p(x_J) \geq \frac{1}{2} \\ 1 + (2p(x_J - 1)) & p(x_J) < \frac{1}{2} \end{cases} \\ &= \begin{cases} 2(1 - p(x_J)) & p(x_J) \geq \frac{1}{2} \\ 2p(x_J) & p(x_J) < \frac{1}{2} \end{cases} \\ &= 2 \min\{p(x_J), 1 - p(x_J)\}. \end{aligned}$$

We obtain

$$\|f - g\|_2^2 = 2^{-|J|} \cdot 4 \sum_{x_J} \min\{p(x_J), 1 - p(x_J)\}.$$

Therefore, we have the following lemma.

Lemma 10. Suppose we have a boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ and we have a fixed $J \subset [n]$. Let g (g_b) be the best real-valued ($\{-1, 1\}$-valued) approximation of f in the L_2 norm, depending only on variables labelled by elements in J. Then

$$\|f - g\|_2^2 = 2^{-|J|} \cdot 4 \sum_{x_J} \min\{p(x_J), 1 - p(x_J)\}$$

and

$$\|f - g_b\|_2^2 = 2^{-|J|} \cdot 4 \sum_{x_J} p(x_J)(1 - p(x_J))\},$$

where $p(x_J) = \mathbb{P}(f = 1|x_J)$. Moreover,

$$\|f - g_b\|_2^2 \leq \|f - g\|_2^2.$$

Proof. We have $\min\{p(x_J), 1 - p(x_J)\} \leq 2p(x_J)(1 - p(x_J))$. □

We prove the following theorem due to E. Friedgut.

Theorem 4 (Friedgut, '98)**.** If $f : \{-1,1\}^n \to \{-1,1\}$ and $I(f) = k$ then for every $\varepsilon > 0$ there exists a boolean function $g : \{-1,1\}^n \to \{-1,1\}$ depending only on $\exp(\lceil ck/\varepsilon \rceil)$ variables, such that $\mathbb{P}(f \neq g) \leq \varepsilon$.

Note that for boolean f, g we have

$$\|f - g\|_2^2 = \mathbb{E}(f - g)^2 = 4\mathbb{P}(f \neq g).$$

Thus it suffices to prove the following theorem

Theorem 5 (Friedgut, '98)**.** If $f : \{-1,1\}^n \to \{-1,1\}$ and $I(f) = k$ then for every $\varepsilon > 0$ there exists a boolean function $g : \{-1,1\}^n \to \{-1,1\}$ depending only on $\exp(\lceil ck/\varepsilon \rceil)$ variables, such that $\|f - g\|_2 \leq \varepsilon$.

Proof. We have seen in the proof of KKL Theorem that if $f_i(x) = f(x) - f(x^i)$ then $\|f_i\|_q^q = 2^q I_i$ and by hypercontractivity

$$\sum_{S:\ i \in S} a_S^2 \delta^{2|S|} \leq I_i^{\frac{2}{1+\delta^2}}.$$

Let

$$J = \{i :\ I_i < \exp(-d)\}.$$

We sum these inequalities for $i \in J$ and we arrive at

$$\sum_S a_S^2 \delta^{2|S|} |S \cap J| \leq \sum_{i \in J} I_i^{\frac{2}{1+\delta^2}}.$$

We obtain

$$\sum_{i \in J} I_i^{\frac{2}{1+\delta^2}} = \sum_{i \in J} I_i \cdot I_i^{\frac{1-\delta^2}{1+\delta^2}} \leq \left(\sum_{i \in J} I_i \right) e^{-d \frac{1-\delta^2}{1+\delta^2}} \leq k e^{-d \frac{1-\delta^2}{1+\delta^2}} = k \exp\left(d \left(1 - \frac{2}{1+\delta^2} \right) \right).$$

We therefore have

$$\sum_S a_S^2 \delta^{2|S|} |S \cap J| \leq k \exp\left(d\left(1 - \frac{2}{1+\delta^2}\right)\right)$$

and

$$\sum_S a_S^2 |S| = k.$$

It follows that

$$\sum_{S\colon\ |S| \geq \frac{4k}{\varepsilon}} a_S^2 \leq \frac{\varepsilon}{4}$$

and

$$\sum_{S\colon\ \delta^{2|S|}|S\cap I| \geq \frac{4k}{\varepsilon}\exp\left(d\left(1-\frac{2}{1+\delta^2}\right)\right)} a_S^2 \leq \frac{\varepsilon}{4}.$$

Therefore almost all of the spectrum in concentrated on S such that

$$|S| < \frac{4k}{\varepsilon}, \qquad \delta^{2|S|}|S \cap I| < \frac{4k}{\varepsilon} \exp\left(d\left(1 - \frac{2}{1+\delta^2}\right)\right).$$

Take such an S and let $M = 4k/\varepsilon$. If $|S \cap I| \neq 0$ then

$$\delta^{2M} < \delta^{2|S|}|S \cap I| < M \exp\left(d\left(1 - \frac{2}{1+\delta^2}\right)\right).$$

Let $x = \delta^2$. We have

$$x^M < M \exp\left(d\left(1 - \frac{2}{1+x}\right)\right).$$

It follows that

$$d < \frac{1+x}{1-x}\left(\ln M - M \ln x\right).$$

Now we optimize the right hand side with respect to $x \in [0, 1]$. We have

$$d < \frac{1+x}{1-x}\left(\ln M - M \ln x\right) \leq M\frac{1+x}{1-x}\left(\frac{\ln M}{M} - \ln x\right) = M\frac{1+x}{1-x}\left(a - \ln x\right),$$

where $a = \frac{\ln M}{M}$. We have

$$\frac{1+x}{1-x}\left(a - \ln x\right) \leq \frac{1+x}{1-x}\left(a - \frac{x-1}{x}\right) = \frac{1+x}{1-x}a + \frac{1+x}{x}.$$

The minimum of the right hand side is attained at $x = \frac{1}{1+\sqrt{2a}}$. We obtain

$$\frac{d}{M} < \frac{1+\frac{1}{1+\sqrt{2a}}}{1-\frac{1}{1+\sqrt{2a}}}a + \frac{1+\frac{1}{1+\sqrt{2a}}}{\frac{1}{1+\sqrt{2a}}} = \frac{2+\sqrt{2a}}{\sqrt{2a}}a + 2 + \sqrt{2a} = (2+\sqrt{2a})(1+\sqrt{a/2})$$

Since $a = \frac{\ln M}{M} \leq \frac{1}{e}$. Therefore

$$\frac{d}{M} < \left(2+\sqrt{2/e}\right)\left(1+\sqrt{1/(2e)}\right) < 5.$$

Thus, if $\frac{d}{M} > 5$, then $|S \cap I| = 0$. Take $d = 5M = \frac{20k}{\varepsilon}$. Therefore, if

$$J = \{i : \; I_i < \exp\left(-\frac{20k}{\varepsilon}\right)\}$$

then

$$\sum_{S:\; |S\cap J|>0} a_S^2 \leq \frac{\varepsilon}{2}.$$

Let us define the function g as follows

$$\hat{g}(S) = \begin{cases} \hat{f}(S) & |S \cap J| = 0 \\ 0 & |S \cap J| \neq 0 \end{cases}.$$

Thus g depends only on the variables in $[n]\backslash J$. We have

$$|[n]\backslash J| e^{-d} \leq k.$$

Therefore

$$|[n]\backslash J| \leq ke^d \leq k \exp\left(\frac{20k}{\varepsilon}\right) \leq \exp\left(\frac{ck}{\varepsilon}\right).$$

Thus

$$\|f-g\|_2^2 = \sum_S (\hat{f}(S) - \hat{g}(S))^2 = \sum_{|S\cap J|>0} a_S^2 \leq \frac{\varepsilon}{2}.$$

$\square$

10 Degree of a boolean function

Let $f : \{-1,1\}^n \to \{-1,1\}$, $f = \sum_S a_S w_S$. We define the *degree* of f by

$$\deg(f) = \max\{0 \leq k \leq n \mid \exists S \; |S| = k, a_S \neq 0\}.$$

In other words, since f is a polynomial in the Walsh representation, the degree of f is simply the degree of this polynomial.

We prove that the boolean function depending on n variables cannot have small degree.

Proposition 7. Suppose $f : \{-1,1\}^n \to \{-1,1\}$ is a boolean function of degree d and suppose that f depends on all of its variables, namely $I_i(f) > 0$ for $i = 1, \ldots, n$. Then

$$n \leq d2^d.$$

Lemma 11. Suppose $f : \{-1,1\}^n \to \mathbb{R}$ and suppose $\deg(f) \leq d$ and f is not identically 0. Then $\mathbb{P}(f \neq 0) \geq 2^{-d}$.

Proof. We prove the lemma by induction on n. For $n = 1$ if $f \equiv c$ then $c \neq 0$ and the statement follows. If f is not constant, then it is a polynomial of degree 1 and $f(x_1) = a + bx_1$ with $b \neq 0$. Therefore, if $f(-1) = a - b = 0$ then $f(1) = a + b \neq 0$ and if $f(1) = a + b = 0$ then $f(-1) = a - b \neq 0$. Therefore always $\mathbb{P}(f \neq 0) \geq \frac{1}{2}$.

Suppose we have $f : \{-1,1\}^n \to \mathbb{R}$, $\deg(f) \leq d$ and f is not identically 0. Let us write f in the form

$$f(x_1, \ldots, x_n) = x_n f_1(x_1, \ldots, x_{n-1}) + f_2(x_1, \ldots, x_{n-1}).$$

Note that $\deg(f_1) \leq d - 1$. If $f_1 - f_2 \equiv 0$ then

$$f(x_1, \ldots, x_n) = (1 + x_n) f_1(x_1, \ldots, x_n).$$

Note that f_1 is not identically 0 since f is not identically 0. By the induction hypothesis we have

$$\mathbb{P}(f \neq 0) = \mathbb{P}(x_n = 1, f_1(x_1, \ldots, x_{n-1}) \neq 0) = \frac{1}{2}\mathbb{P}(f_1 \neq 0) \geq \frac{1}{2} \cdot 2^{-(d-1)} = 2^{-d}.$$

In the same way we treat the case when $f_1 + f_2 \equiv 0$.

Now suppose that $f_1 - f_2$ and $f_1 + f_2$ are not identically 0. Clearly $\deg(f_1 - f_2) \leq d$ and $\deg(f_1 + f_2) \leq d$. Therefore,

$$\mathbb{P}(f \neq 0) = \mathbb{P}(f_1 - f_2 \neq 0, x_n = -1) + \mathbb{P}(f_1 + f_2 \neq 0, x_n = 1) \geq \frac{1}{2}2^{-d} + \frac{1}{2}2^{-d} = 2^{-d}.$$

□

Proof of Proposition 7. Suppose $f : \{-1,1\}^n \to \{-1,1\}$ satisfies $\deg(f) \le d$. Take $f_i(x) = f(x) - f(x^i)$. Since $I_i(f) > 0$ we have that f_i is not identically 0. Therefore, from the lemma we have

$$I_i(f) = \mathbb{P}(f_i \neq 0) \ge 2^{-d}.$$

Thus

$$n2^{-d} \le \sum_{i=1}^{n} I_i(f) = I(f) = \sum_S a_S^2 |S| \le d \sum_S a_S^2 = d.$$

Thus $n \le d2^d$. $\square$

Now we prove a proposition about the algebraic properties of a spectrum of a function $f : \{-1,1\}^n \to \mathbb{Z}$.

Proposition 8. Suppose $f : \{-1,1\}^n \to \mathbb{Z}$ satisfies $\deg(f) \le d$. Then $\hat{f}(S) = a(S)2^{-d}$, where $a(s) \in \mathbb{Z}$.

Proof. Induction on d. If $d = 0$ then the assertion is trivial. Take $f_i(x) = f(x) - f(x^i)$. Then

$$f_i = 2 \sum_{S \subset [n] \setminus \{i\}} \hat{f}(S \cup \{i\}) w_{S \cup \{i\}}.$$

Clearly,

$$x_i f_i(x) = 2 \sum_{S \subset [n] \setminus \{i\}} \hat{f}(S \cup \{i\}) w_S(x)$$

and this function has degree at most $d - 1$. Thus $2\hat{f}(S \cup \{i\}) = a(S)2^{-(d-1)}$. We obtain $\hat{f}(S \cup \{i\}) = a(S)2^{-d}$. Since every nonempty set $S \subset [n]$ can we written in the form $S = S' \cup \{i\}$ for some i, our assertion follows for this sets. We also have $\hat{f}(\emptyset) = a(\emptyset)2^{-d}$. Indeed,

$$\hat{f}(\emptyset) = f - \sum_{S \neq \emptyset} a(S)2^{-d} w_S.$$

The right hand side clearly is a number in $2^{-d}\mathbb{Z}$. $\square$

Note that from the above statement it follows that for every boolean function $f : \{-1,1\}^n \to \{-1,1\}$ with $I_i(f) > 0$ for all $1 \le i \le n$ we have $n \le d4^d$. Indeed, we have

$$I_i(f) = \sum_{S:\ i \in S} \hat{f}(S)^2 \ge (2^{-d})2 = 4^{-d}.$$

Thus

$$n4^{-d} \le \sum_{i=1}^{n} I_i(f) = I(f) \le d.$$

Recall now the general statement of the hypercontractivity.

Theorem 6. Let $p \geq q > 1$. Then for $0 \leq \delta \leq \sqrt{\frac{q-1}{p-1}}$ we have

$$\|T_\delta f\|_p \leq \|f\|_q.$$

Using this theorem we prove the following proposition.

Proposition 9. Let $f : \{-1,1\}^n \to \{-1,1\}$ and $\deg(f) \leq d$. Then for $q \in [1,2]$ we have

$$(q-1)^{d/2}\|f\|_2 \leq \|f\|_q.$$

Moreover, if $p \geq 2$ then

$$\|f\|_p \leq (p-1)^{d/2}\sqrt{d+1}\|f\|_2.$$

Proof. Take $p = 2$ and $\delta = \sqrt{q-1}$. We then have

$$(q-1)^d\|f\|_2^2 = \delta^{2d}\sum_S a_S^2 \leq \sum_S \delta^{2|S|}a_S^2 = \|T_\delta f\|_2^2 \leq \|f\|_q^2.$$

To prove the second part let us take $q = 2$ and $\delta = \frac{1}{\sqrt{p-1}}$, $p \geq 2$. Let

$$f_k = \sum_{S:\ |S|=k} a_S w_S.$$

Then

$$(p-1)^{-d/2}\|f_k\|_p \leq (p-1)^{-k/2}\|f_k\|_p = \delta^k\|f_k\|_p = \|\sum_{S:\ |S|=k}\delta^k a_S w_S\|_p$$
$$= \|T_\delta f\|_p \leq \|f_k\|_2.$$

Thus,

$$\|f_k\|_p \leq (p-1)^{d/2}\|f_k\|_2.$$

Therefore,

$$\|f\|_p \leq \sum_{k=0}^n \|f_k\|_p \leq (p-1)^{d/2}\sum_{k=0}^d \|f_k\|_2 \leq (p-1)^{d/2}\sqrt{d+1}\sqrt{\sum_{k=0}^d \|f_k\|_2^2}$$
$$= (p-1)^{d/2}\sqrt{d+1}\|f\|_2,$$

since $(f_k)_{k=0,1,\dots,d}$ are orthogonal. $\square$

Proposition 10. Let $f : \{-1,1\}^n \to \{-1,0,1\}$. Then for every $\delta \in [0,1]$ we have

$$\sum_S \delta^{|S|} a_S^2 \leq \mathbb{P}(f \neq 0)^{\frac{2}{1+\delta}}.$$

Proof. We have

$$\sum_S \delta^{|S|} a_S^2 = \|T_{\sqrt{\delta}} f\|_2^2 \leq \|f\|_{1+\delta}^2 = \mathbb{P}(f \neq 0)^{\frac{2}{1+\delta}}.$$

□

Note that since $\sum_S a_S^2 = \mathbb{P}(f \neq 0)$ then for f not identically 0 we have

$$\frac{\sum_S \delta^{|S|} a_S^2}{\sum_S a_S^2} \leq \mathbb{P}(f \neq 0)^{\frac{1-\delta}{1+\delta}}.$$

Therefore, if f has small support, then the spectrum of f cannot be concentrated on the low-degree Fourier levels. It also follows that

$$\delta^d \leq |\text{supp } f|^{\frac{1-\delta}{1+\delta}}.$$

Therefore, the $\{-1,0,1\}$-valued boolean function with a very small support must have large degree.